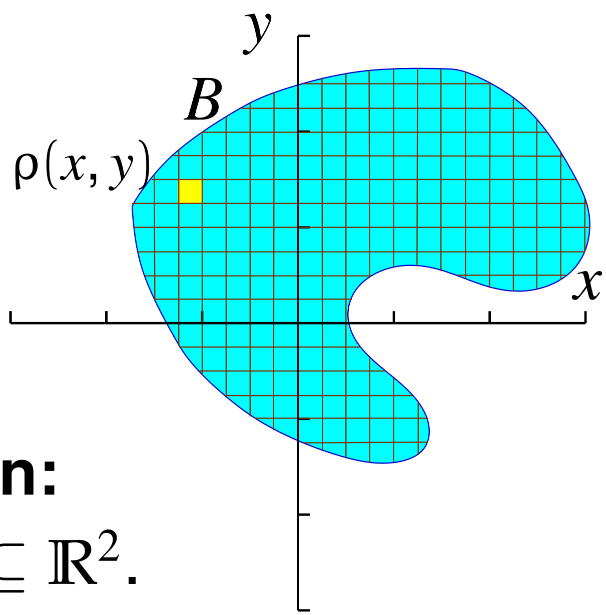
3.7, 3.8, 3.9, 3.11 Functions of multiple random variables (continuous)

Prof. Tesler

Math 186
Winter 2019

Mass density (review from Calculus and Physics)



Two-dimensional version:

- Consider a shape $B \subseteq \mathbb{R}^2$.
- Make very thin horizontal and vertical cuts.
- Let $\rho(x, y)$ be the *density* at (x, y). This is the mass per unit area.
- It can be measured in g/cm^2. In 3D, it would be g/cm^3.
- $\rho(x, y) \geqslant 0$ everywhere.
- The area of a differential patch is $dA = dx\,dy = dy\,dx$.
- The mass of a differential patch is $\rho(x, y)\,dA$ (density times area).
- The total mass of B is $\displaystyle\iint_B \rho(x, y)\,dA$

Continuous joint probability density function

Joint probability density function of two variables

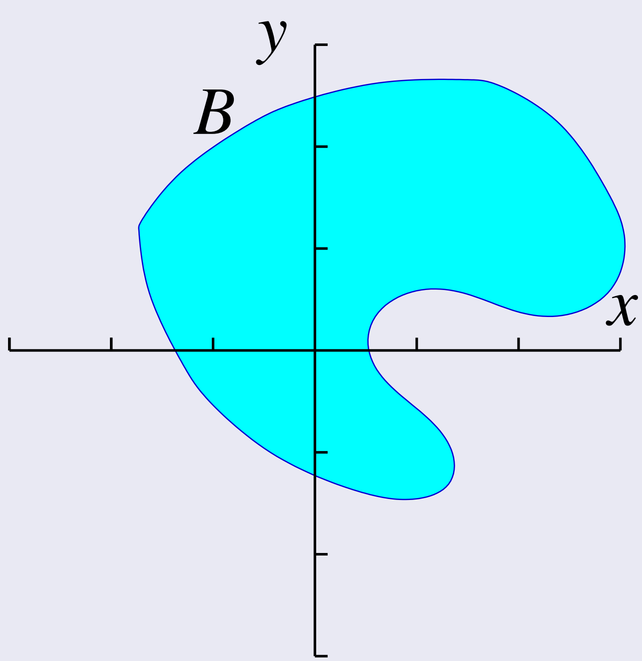
We require:

- $f_{X,Y}(x,y) \geqslant 0$ for all points (x,y).
- $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$

Probability of an event

The probability of event $B \subseteq \mathbb{R}^2$ is

$$P(B) = \iint_B f_{X,Y}(x,y)\,dA$$



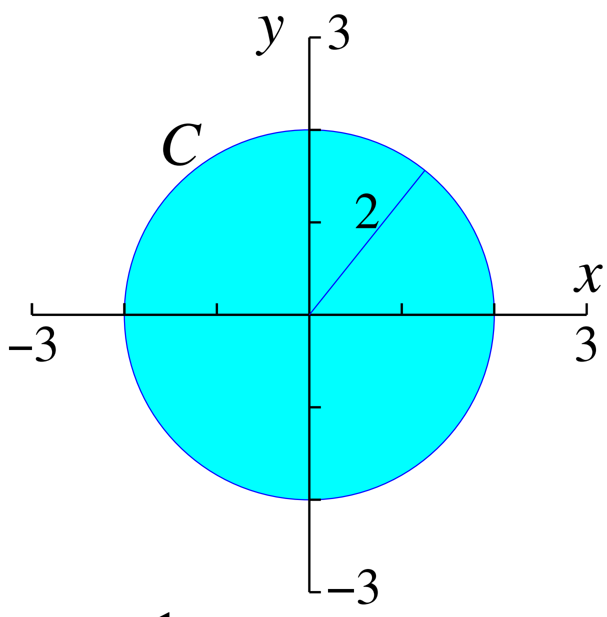
Uniform probability on a region C

- *Uniform probability* on a region C means that all points inside C have equal probability density, and all points outside C have probability density 0:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\text{area}(C)} & \text{if } (x,y) \in C \\ 0 & \text{otherwise} \end{cases}$$

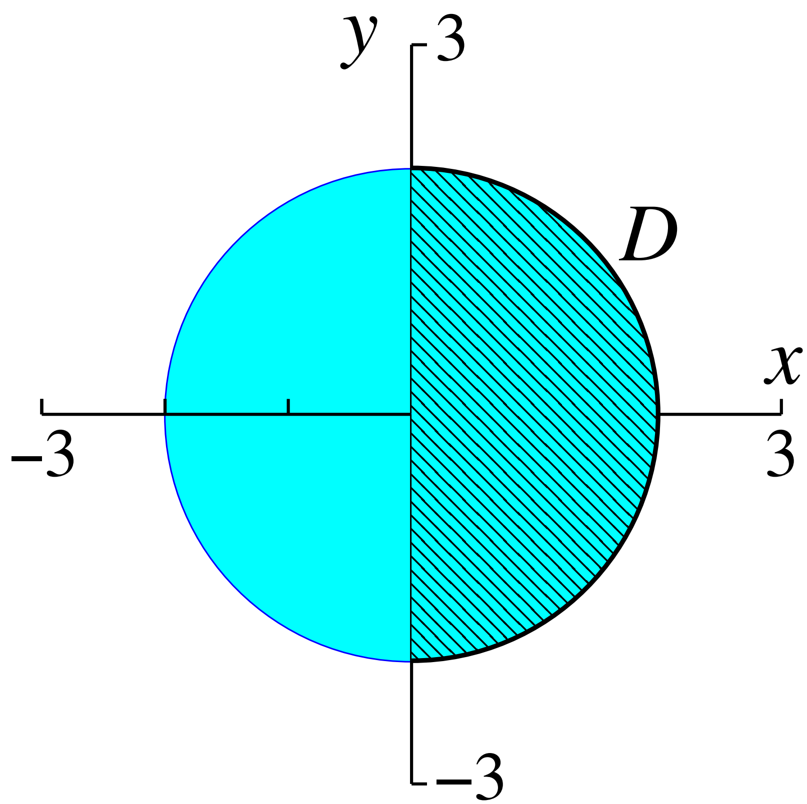
- Let C be the disk of radius 2 centered at the origin:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4\pi} & \text{if } x^2 + y^2 \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$



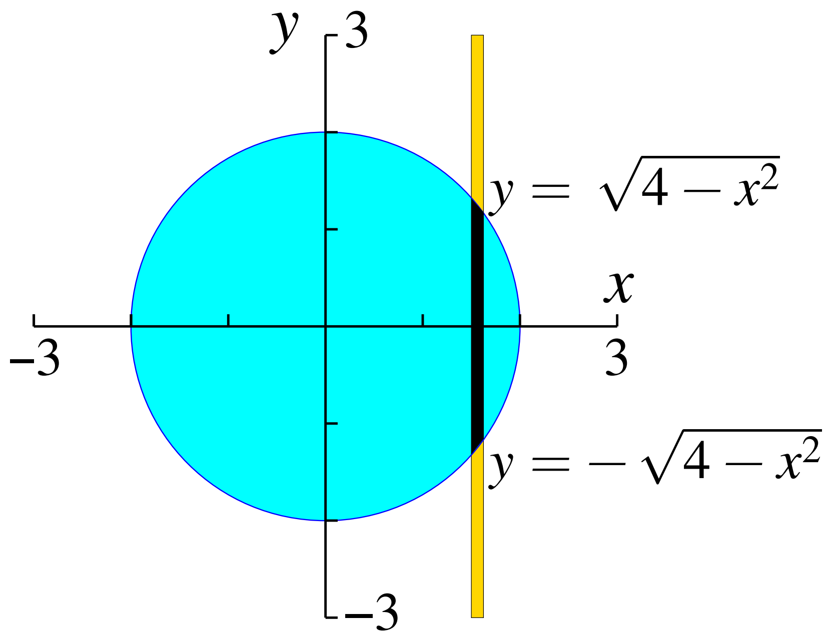
- Total probability = $\iint_C \frac{1}{4\pi}\, dA = \frac{1}{4\pi} \cdot \text{area}(C) = \frac{1}{4\pi} \cdot 4\pi = 1$

Probability of an event



$$P(X > 0) = \iint_D \frac{1}{4\pi}\,dA = \frac{1}{4\pi}\text{area}(D) = \frac{1}{4\pi}\cdot\frac{4\pi}{2} = \boxed{\mathbf{\frac{1}{2}}}$$

Marginal densities



- Form an x-strip: hold x constant and vary y.
- The perimeter is $x^2 + y^2 = 4$, so $y = \pm\sqrt{4-x^2}$ on the perimeter.
- The strip is vertical, so the $-$ solution is at the bottom and the $+$ solution is at the top.
- The part of the strip within the shape is $-\sqrt{4-x^2} \leqslant y \leqslant \sqrt{4-x^2}$.

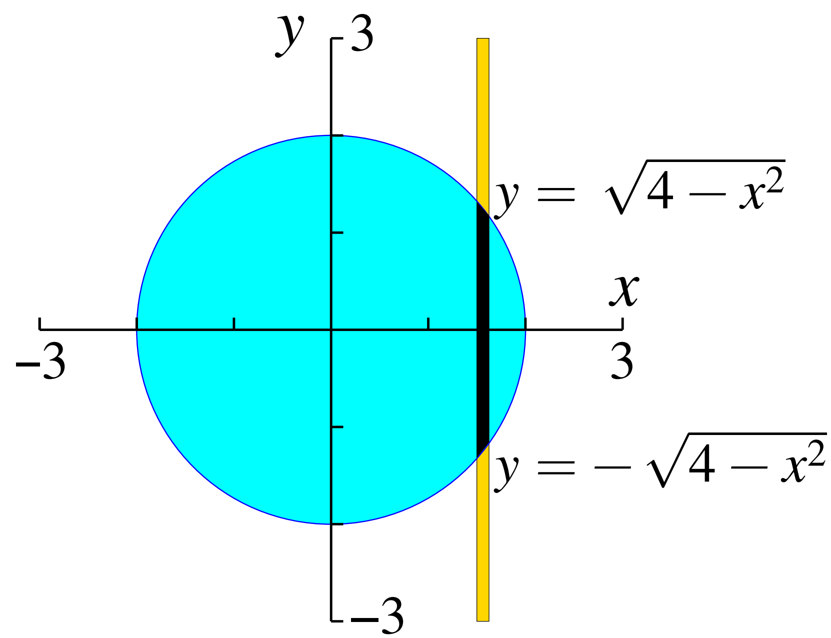
Marginal densities

The marginal density at x:

Form an *x-strip*.

Hold x constant and integrate over all y.

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \\
&= \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \frac{1}{4\pi}\,dy \\
&= \frac{2\sqrt{4-x^2}}{4\pi}
\end{aligned}
$$



$$
f_X(x) = \begin{cases} \frac{\sqrt{4-x^2}}{2\pi} & \text{if } -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}
$$

Marginal densities

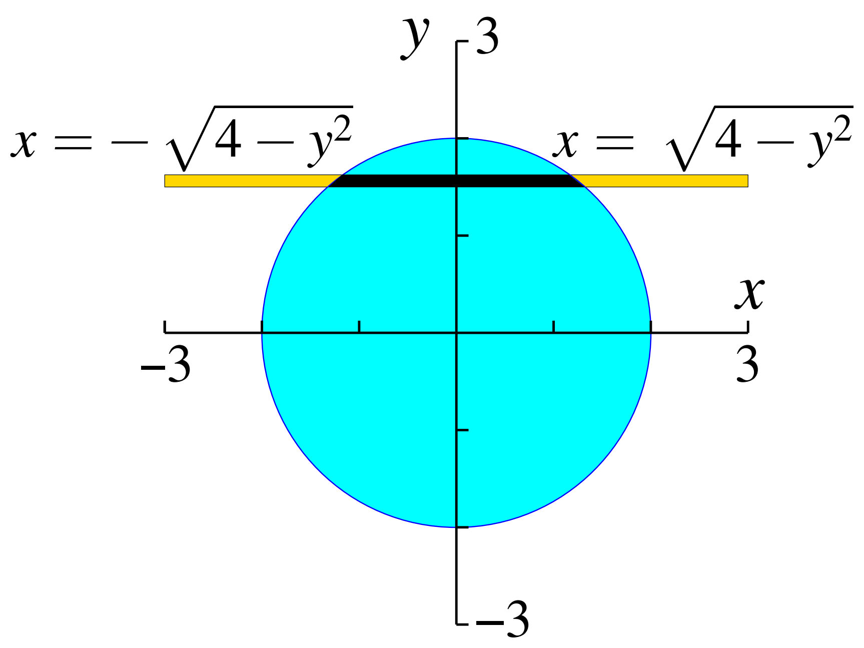
The marginal density at y is similar.

Form a *y-strip*.

Hold y constant and integrate over all x.

The strip is horizontal and goes left to right instead of bottom to top.

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \\ &= \int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} \frac{1}{4\pi}\,dx \\ &= \frac{2\sqrt{4-y^2}}{4\pi} \end{aligned}$$



$$f_Y(y) = \begin{cases} \frac{\sqrt{4-y^2}}{2\pi} & \text{if } -2 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Independence

- Random variables $X, Y, Z, \ldots$ are *independent* if their joint pdf factorizes as follows, for all $x, y, z, \ldots$.

$$f_{X,Y,Z,\ldots}(x, y, z, \ldots) = f_X(x)\, f_Y(y)\, f_Z(z) \cdots$$

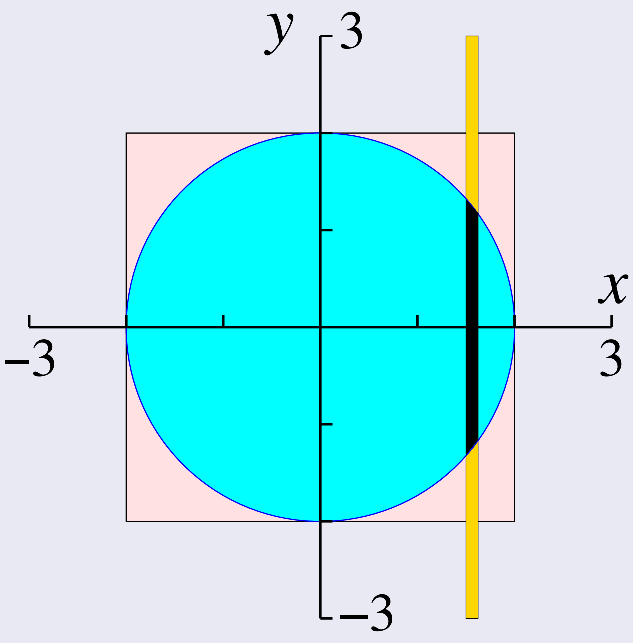
- **Technicality:** Exceptions are allowed, as long as the probability of an exception is 0. For example, in a continuous distribution:
  - The probability of a point is 0.
  - In 2D, the probability of a discrete set of points or curves is 0.
  - Exceptions only happen with continuous distributions, not discrete.

Summary of previous formulas

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4\pi} & \text{if } x^2 + y^2 \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = \begin{cases} \sqrt{4 - x^2}/(2\pi) & \text{if } -2 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \sqrt{4 - y^2}/(2\pi) & \text{if } -2 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$



Check independence:

$$f_X(x)f_Y(y) = \begin{cases} \sqrt{(4 - x^2)(4 - y^2)}/(4\pi^2) & \text{if } -2 \leqslant x \leqslant 2 \text{ and } -2 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

This is different than $f_{X,Y}(x,y)$. The formula is different, and it's nonzero inside a square instead of inside a circle. So X, Y are dependent.

Expected values

Definition

For a function $g(X,Y)$ of continuous random variables, the *expected value* is

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dA$$

This is similar to the definition in the discrete case, but using integrals instead of sums.

Compute $E(X)$ for the circle example

$$E(X) = \iint\limits_{\text{left semicircle}} \frac{x}{4\pi}\,dA + \iint\limits_{\text{right semicircle}} \frac{x}{4\pi}\,dA = \boxed{\mathbf{0}}$$

The two integrals are negatives of each other, so they sum to 0.

Compute $E(R)$ in the circle example

In polar coordinates, recall $R = \sqrt{X^2 + Y^2}$. Compute $E(R)$:

$$E(R) = E\left(\sqrt{X^2+Y^2}\right) = \iint_C \sqrt{x^2+y^2} \cdot \frac{1}{4\pi}\, dA$$

- This is easier in polar coordinates than in Cartesian coordinates. Switch to polar coordinates, and note that the integral separates:

$$E(R) = \iint_C \frac{r}{4\pi}\, dA = \int_0^{2\pi}\int_0^2 \frac{r}{4\pi} \cdot r\, dr\, d\theta = \frac{1}{4\pi}\left(\int_0^{2\pi} d\theta\right)\left(\int_0^2 r^2\, dr\right)$$

- Evaluate the integrals:

$$\int_0^{2\pi} d\theta = \theta\Big|_{\theta=0}^{\theta=2\pi} = 2\pi - 0 = 2\pi \qquad \int_0^2 r^2\, dr = \frac{r^3}{3}\Big|_{r=0}^{r=2} = \frac{2^3-0^3}{3} = \frac{8}{3}$$

- Plug in their values:

$$E(R) = \frac{1}{4\pi}(2\pi)\left(\frac{8}{3}\right) = \frac{16\pi}{12\pi} = \boxed{\mathbf{\frac{4}{3}}}$$

Variance

- The variance formula is the same for continuous as for discrete:
$$\mathrm{Var}(X) = E((X-\mu)^2) = E(X^2) - (E(X))^2$$
However, expected value is computed using an integral instead of a sum.
- Compute $\mathrm{Var}(R)$ and $\mathrm{SD}(R)$:

$$E(R^2) = \iint_C \frac{r^2}{4\pi}\,dA = \int_0^{2\pi}\int_0^2 \frac{r^2}{4\pi}\cdot r\,dr\,d\theta = \frac{1}{4\pi}\left(\int_0^{2\pi} d\theta\right)\left(\int_0^2 r^3\,dr\right)$$

$$\int_0^{2\pi} d\theta = 2\pi \qquad \int_0^2 r^3\,dr = \left.\frac{r^4}{4}\right|_{r=0}^{2} = \frac{2^4-0^4}{4} = 4$$

$$E(R^2) = \frac{1}{4\pi}(2\pi)(4) = 2$$

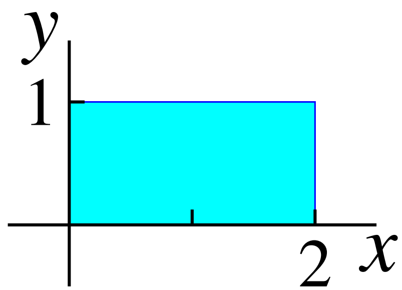
$$\mathrm{Var}(R) = E(R^2) - (E(R))^2 = 2 - (4/3)^2 = \boxed{\mathbf{2/9}} \qquad \mathrm{SD}(R) = \sqrt{2/9}$$

Mass density in physics vs. continuous pdf

Physics	Probability
Mass density $\rho(x,y) \geqslant 0$	**Probability density function** $f_{X,Y}(x,y) \geqslant 0$
Mass of shape $D \subseteq \mathbb{R}^2$: $M = \iint_D \rho(x,y)\,dA \geqslant 0$	**Probability of event $D \subseteq \mathbb{R}^2$:** $P(D) = \iint_D f_{X,Y}(x,y)\,dA$ and $P(\mathbb{R}^2) = 1$
Center of mass $(\bar{x},\bar{y})$ $\bar{x} = \dfrac{\iint_D x \cdot \rho(x,y)\,dA}{\iint_D \rho(x,y)\,dA}$ $\bar{y}$ formula is similar	**Expected value** $E(X) = \iint_{\mathbb{R}^2} x \cdot f_{X,Y}(x,y)\,dA$ = numerator of $\bar{x}$. The denominator of $\bar{x}$ is 1, since $M = 1$. $E(Y)$ formula is similar

- When the total mass is 1, we have $(\bar{x},\bar{y}) = (E(X),E(Y))$.
- We used $\mathbb{R}^2$. For $\mathbb{R}^n$, use $(x_1,\ldots,x_n)$ instead of (x,y).

Determining the constant



Question: Determine the formula of the probability density if it is proportional to $x + 4y$ inside the rectangle and is 0 outside.

- We have $f_{X,Y}(x, y) = c(x + 4y)$ inside the rectangle and 0 outside, for some constant c.
- Find c so that the total probability is 1:

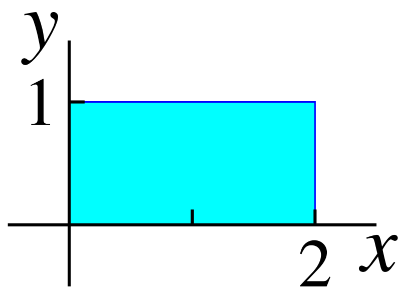
$$P = \int_0^2 \int_0^1 c(x + 4y)\, dy\, dx = 1$$

- The inside integral is

$$c(xy + 2y^2)\Big|_{y=0}^{y=1} = c(x(1-0) + 2(1^2 - 0^2)) = c(x+2)$$

- Plug that back in: $P = \int_0^2 c(x + 2)\, dx$

Determining the constant



- Continue evaluating:

$$P = \int_0^2 c(x+2)\,dx = c\left(\frac{x^2}{2}+2x\right)\Big|_{x=0}^{2}$$
$$= c\left(\frac{2^2-0^2}{2}+2(2-0)\right) = c\cdot(2+4) = 6c$$

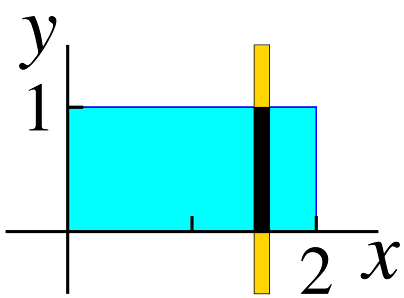
- To get $P=1$, solve $6c=1$, so $c=1/6$.
- Plug this value of c into the formula $f_{X,Y}(x,y)=c(x+4y)$.
- Thus, the pdf is

$$f_{X,Y}(x,y) = \begin{cases} \frac{x+4y}{6} & \text{inside rectangle: } 0 \leqslant x \leqslant 2 \text{ and } 0 \leqslant y \leqslant 1 \\ 0 & \text{outside rectangle} \end{cases}$$

Marginal densities for rectangle example

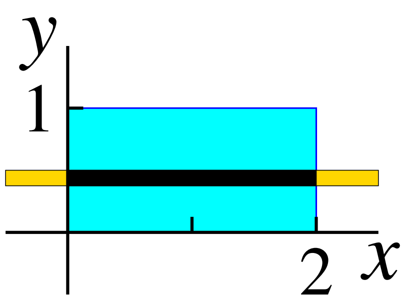
- For $0 \leqslant x \leqslant 2$:

$$f_X(x) = \int_0^1 \frac{x+4y}{6}\,dy = \left.\frac{xy+2y^2}{6}\right|_{y=0}^{1} = \frac{x(1-0)+2(1^2-0^2)}{6} = \frac{x+2}{6}$$



- Otherwise, $f_X(x) = 0$.

Marginal densities for rectangle example



- For $0 \leqslant y \leqslant 1$:

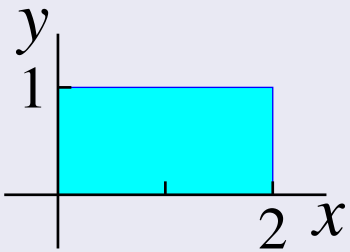
$$
\begin{aligned}
f_Y(y) &= \int_0^2 \frac{x+4y}{6}\,dx = \left(\frac{x^2}{12} + \frac{4xy}{6}\right)\Big|_{x=0}^{2} \\
&= \left(\frac{2^2-0^2}{12} + \frac{4(2-0)y}{6}\right) \\
&= \frac{4}{12} + \frac{8y}{6} = \frac{4y+1}{3}
\end{aligned}
$$

- Otherwise, $f_Y(y) = 0$.

Independence in rectangle example

Summary of formulas

$$f_{X,Y}(x,y) = \begin{cases} \frac{x+4y}{6} & \text{inside rectangle: } 0 \leqslant x \leqslant 2 \text{ and } 0 \leqslant y \leqslant 1 \\ 0 & \text{outside rectangle} \end{cases}$$



$$f_X(x) = \begin{cases} \frac{x+2}{6} & \text{if } 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{4y+1}{3} & \text{if } 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

Check independence

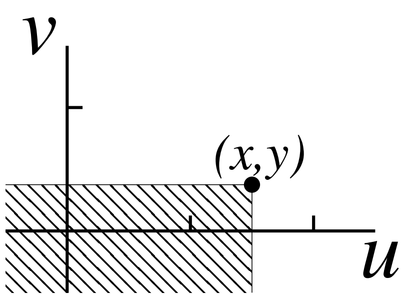
$$f_X(x) \cdot f_Y(y) = \begin{cases} \frac{(x+2)(4y+1)}{18} & \text{if } 0 \leqslant x \leqslant 2 \text{ and } 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$\neq f_{X,Y}(x,y)$, so X and Y are *dependent*.

Joint Cumulative Distribution Function (cdf)

- The *joint cumulative distribution function (cdf)* for two random variables X, Y is
$$F_{X,Y}(x,y) = P(X \leqslant x, Y \leqslant y)\,.$$
- For multiple random variables, the formula is similar.
- As an integral:

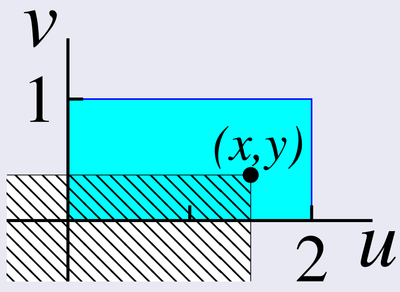
$$\begin{aligned} F_{X,Y}(x,y) &= P(X \leqslant x, Y \leqslant y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du \\ &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\,du\,dv \end{aligned}$$



- Since we used (x,y) in the limits of the integral, the integration variables had to be renamed; here, we used (u,v) instead. Alternatively, some people prefer to do it the other way around: $F_{X,Y}(u,v) = P(X \leqslant u, Y \leqslant v) = \int_{-\infty}^{v} \int_{-\infty}^{u} f_{X,Y}(x,y)\,dx\,dy$.

Joint cdf in rectangle example

$$f_{X,Y}(x,y) = \begin{cases} \frac{x+4y}{6} & \text{inside rectangle:} \\ & 0 \leqslant x \leqslant 2 \text{ and } 0 \leqslant y \leqslant 1 \\ 0 & \text{outside rectangle} \end{cases}$$



- Consider (x,y) inside the rectangle:

$$F_{X,Y}(x,y) = \int_0^x \int_0^y \frac{u+4v}{6}\,dv\,du$$

- Inside integral:

$$\int_0^y \frac{u+4v}{6}\,dv = \left.\frac{uv+2v^2}{6}\right|_{v=0}^{v=y} = \frac{u(y-0)+2(y^2-0^2)}{6} = \frac{uy+2y^2}{6}$$

- Outside integral:

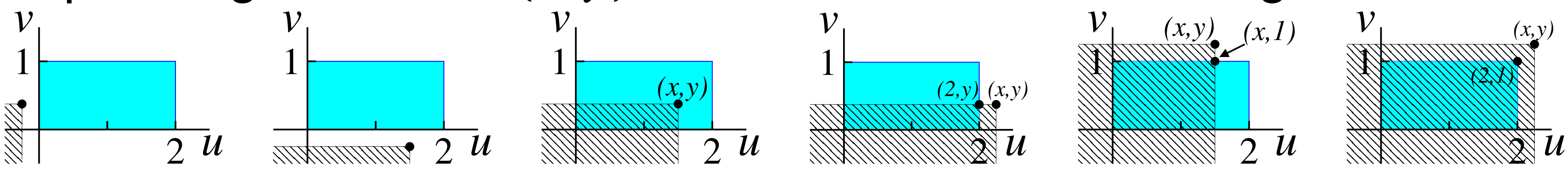
$$F_{X,Y}(x,y) = \int_0^x \frac{uy+2y^2}{6}\,du = \left.\left(\frac{u^2y}{12} + \frac{uy^2}{3}\right)\right|_{u=0}^{u=x} = \frac{x^2y}{12} + \frac{xy^2}{3}$$

Differentiating the cdf

- Evaluate
$$\frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y} F_{X,Y}(x,y)\right)$$
- Inside derivative:
$$\frac{\partial}{\partial y} F_{X,Y}(x,y) = \frac{\partial}{\partial y}\left(\frac{x^2 y}{12} + \frac{x y^2}{3}\right) = \frac{x^2}{12} + \frac{2xy}{3}$$
- Outside derivative:
$$\frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y) = \frac{\partial}{\partial x}\left(\frac{x^2}{12} + \frac{2xy}{3}\right) = \frac{x}{6} + \frac{2y}{3} = \frac{x+4y}{6} = f_{X,Y}(x,y)$$
- In general, the cdf is the double integral of the pdf with respect to x and y, and inversely, the pdf is the double derivative of the cdf with respect to x and y.

Joint cdf in rectangle example

- The pdf $f_{X,Y}(x,y) = (x+4y)/6$ is nonzero only in the blue rectangle.
- The region $\leqslant (x,y)$ can intercept the rectangle in different ways, depending on where (x,y) is in relation to the rectangle.

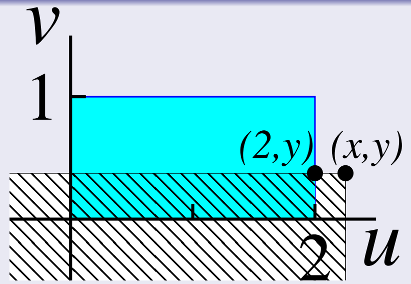


- If $x<0$ or $y<0$, then the pdf is 0 in the whole integration region, so $F_{X,Y}(x,y)=0$.
  This satisfies $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y}F_{X,Y}(x,y) = 0$.

Remaining cases: In this example, when (x, y) is right of and/or above the rectangle, the intercepted region becomes the cdf of another point:

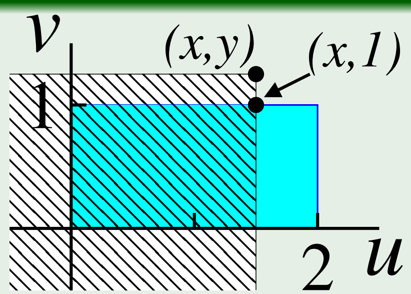
Right: If $x > 2$ and $0 \leqslant y \leqslant 1$

$$F_{X,Y}(x, y) = F_{X,Y}(2, y) = \frac{4y}{12} + \frac{2y^2}{3} = \frac{y + 2y^2}{3}$$



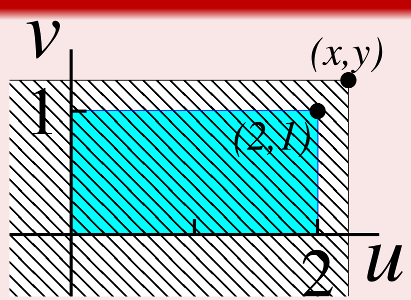
Above: If $y > 1$ and $0 \leqslant x \leqslant 2$

$$F_{X,Y}(x, y) = F_{X,Y}(x, 1) = \frac{x^2}{12} + \frac{x}{3} = \frac{x^2 + 4x}{12}$$



Above and right: If $x > 2$ and $y > 1$

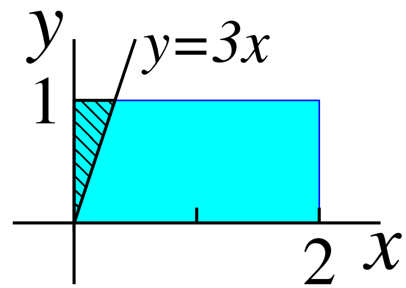
$$F_{X,Y}(x, y) = F_{X,Y}(2, 1) = \frac{2^2 \cdot 1}{12} + \frac{2 \cdot 1}{3} = 1$$



In all of these, $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) = 0$.

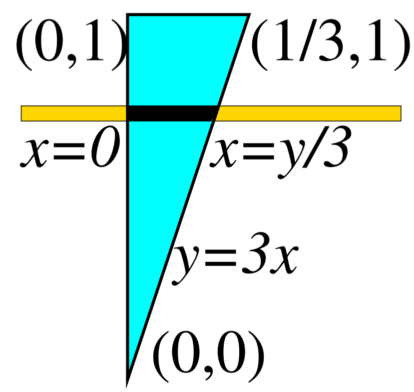
Probability of an event

Evaluate $P(Y > 3X)$ in the rectangle example:



- Compute $\iint_D f_{X,Y}(x,y)\,dA = \iint_D \frac{x+4y}{6}\,dA$ over the shaded triangle, D.
- Can use x-slices or y-slices. Both give the same final answer. x-slices are left as an exercize for you. The y-slices are:

One y-slice for each $0 \leqslant y \leqslant 1$.
It runs over $0 \leqslant x \leqslant y/3$.



- The integral is

$$P(Y > 3X) = \int_0^1 \int_0^{y/3} \frac{x+4y}{6}\,dx\,dy$$

Probability of an event

$$P(Y > 3X) = \int_0^1 \int_0^{y/3} \frac{x+4y}{6}\,dx\,dy$$

- Inside integral:

$$\int_0^{y/3} \frac{x+4y}{6}\,dx \quad = \left(\frac{x^2}{12} + \frac{4xy}{6}\right)\Bigg|_{x=0}^{x=y/3}$$

$$= \frac{(y/3)^2 - 0^2}{12} + \frac{4(y/3)y}{6} = \frac{y^2}{108} + \frac{4y^2}{18} = \frac{25y^2}{108}$$

- Outside integral:

$$P(Y > 3X) = \int_0^1 \frac{25y^2}{108}\,dy = \frac{25y^3}{324}\Bigg|_0^1 = \frac{25(1^3 - 0^3)}{324} = \boxed{\mathbf{\frac{25}{324}}}$$

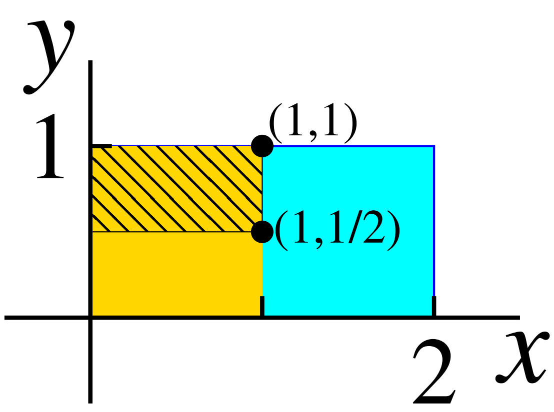
Conditional probability example #1

$P(A|B)$ where A and B have the same dimension as sample space (2D in this example)

Evaluate $P(Y > \frac{1}{2} \mid X < 1)$ in the rectangle example:

- This is $1 - P(Y \leqslant \frac{1}{2} \mid X < 1)$. We have:

$$P(Y \leqslant \tfrac{1}{2} \mid X < 1) = \frac{P(Y \leqslant \frac{1}{2} \text{ and } X < 1)}{P(X < 1)} = \frac{F_{X,Y}(1, 1/2)}{F_X(1)}$$



- Recall that inside the rectangle, we have

$$F_{X,Y}(x, y) = \frac{x^2 y}{12} + \frac{x y^2}{3}$$

and we used tricks to evaluate it outside the rectangle.

- $F_{X,Y}(1, 1/2) = \frac{(1^2)(1/2)}{12} + \frac{(1)(1/2)^2}{3} = \frac{1}{24} + \frac{1}{12} = \frac{3}{24} = \frac{1}{8}$
- $F_X(1) = F_{X,Y}(1, \infty) = F_{X,Y}(1, 1) = \frac{(1^2)(1)}{12} + \frac{(1)(1^2)}{3} = \frac{1}{12} + \frac{1}{3} = \frac{5}{12}$
- Plug these into the above formulas to get

$$\begin{aligned} P(Y > \tfrac{1}{2} \mid X < 1) &= 1 - P(Y \leqslant \tfrac{1}{2} \mid X < 1) = 1 - \frac{F_{X,Y}(1,1/2)}{F_X(1)} \\ &= 1 - \frac{1/8}{5/12} = 1 - \frac{3}{10} = \frac{7}{10} \end{aligned}$$

Conditional probability example #2

$P(A|B)$ where B has smaller dimension than the sample space

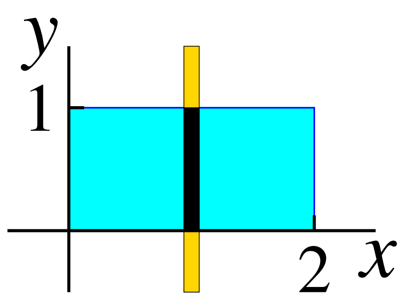
Evaluate $P(Y > \frac{1}{2} \mid X = 1)$ in the rectangle example:

- $\frac{P\left(Y>\frac{1}{2} \text{ and } X=1\right)}{P(X=1)} = \frac{0}{0}$ does not work.
- In $P(A|B) = \frac{P(A \cap B)}{P(B)}$, both $A \cap B$ and B are 1D subspaces of a 2D sample space, so their probabilities are 0.
- Redo this as ratio of probabilities in 1D.

Conditional probability example #2

$P(A|B)$ where B has smaller dimension than the sample space

- Define the *conditional probability density* at $X = x$:



$$f_Y(y \mid X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

- For a given value of x, this is a function of varying y.
- It's proportional to $f_{X,Y}(x,y)$ but is renormalized so that the total probability as y varies in the strip $X = x$ is 1:

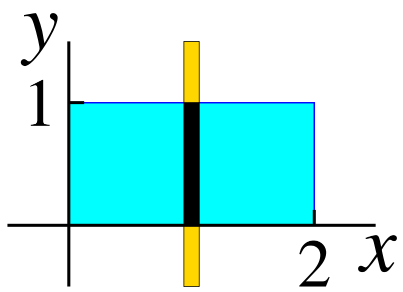
$$\int_{-\infty}^{\infty} f_Y(y \mid X = x)\, dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}\, dy = \frac{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$$

Conditional probability example #2

$P(A|B)$ where B has smaller dimension than the sample space

Evaluate $P(Y > \frac{1}{2} \mid X = 1)$ in the rectangle example:

- The conditional probability density at $X = x$ is



$$f_Y(y \mid X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

- In the rectangle example, for x and y within the rectangle:

$$f_Y(y \mid X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{(x+4y)/6}{(x+2)/6} = \frac{x+4y}{x+2}$$

$$f_Y(y \mid X = 1) = \frac{1+4y}{3}$$

$$P(Y > \tfrac{1}{2} \mid X = 1) = \int_{1/2}^{\infty} f_Y(y \mid X = 1)\,dy = \int_{1/2}^{1} \frac{1+4y}{3}\,dy = \left.\frac{y+2y^2}{3}\right|_{y=1/2}^{y=1}$$

$$= \frac{1-(1/2)}{3} + \frac{2(1^2-(1/2)^2)}{3} = \frac{1/2}{3} + \frac{2(3/4)}{3} = \boxed{\mathbf{\frac{2}{3}}}$$