Lecture 9:

Joint distributions

Joint, marginal and conditional

Joint, marginal and conditional

The distribution of a single r.v. X gives info about the probability of X falling into any subset of the real line.

The **joint** distribution of two r.v.-s X and Y gives complete info about the probability of the vector (X, Y) falling into any subset of the plane.

The **marginal** distribution is the individual distribution of X, ignoring the value of Y,
and the **conditional** distribution of X given $Y = y$ is the updated distribution for X after observing $Y = y$.

Joint, marginal and conditional:

Discrete case

Joint, marginal and conditional: discrete

Definition 7.1.1 (Joint CDF). The **joint** CDF of r.v.-s X and Y is the function $F_{X,Y}$ given by $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$. (For n r.v.-s it is defined analogously. This works for continuous r.v.-s as well!).

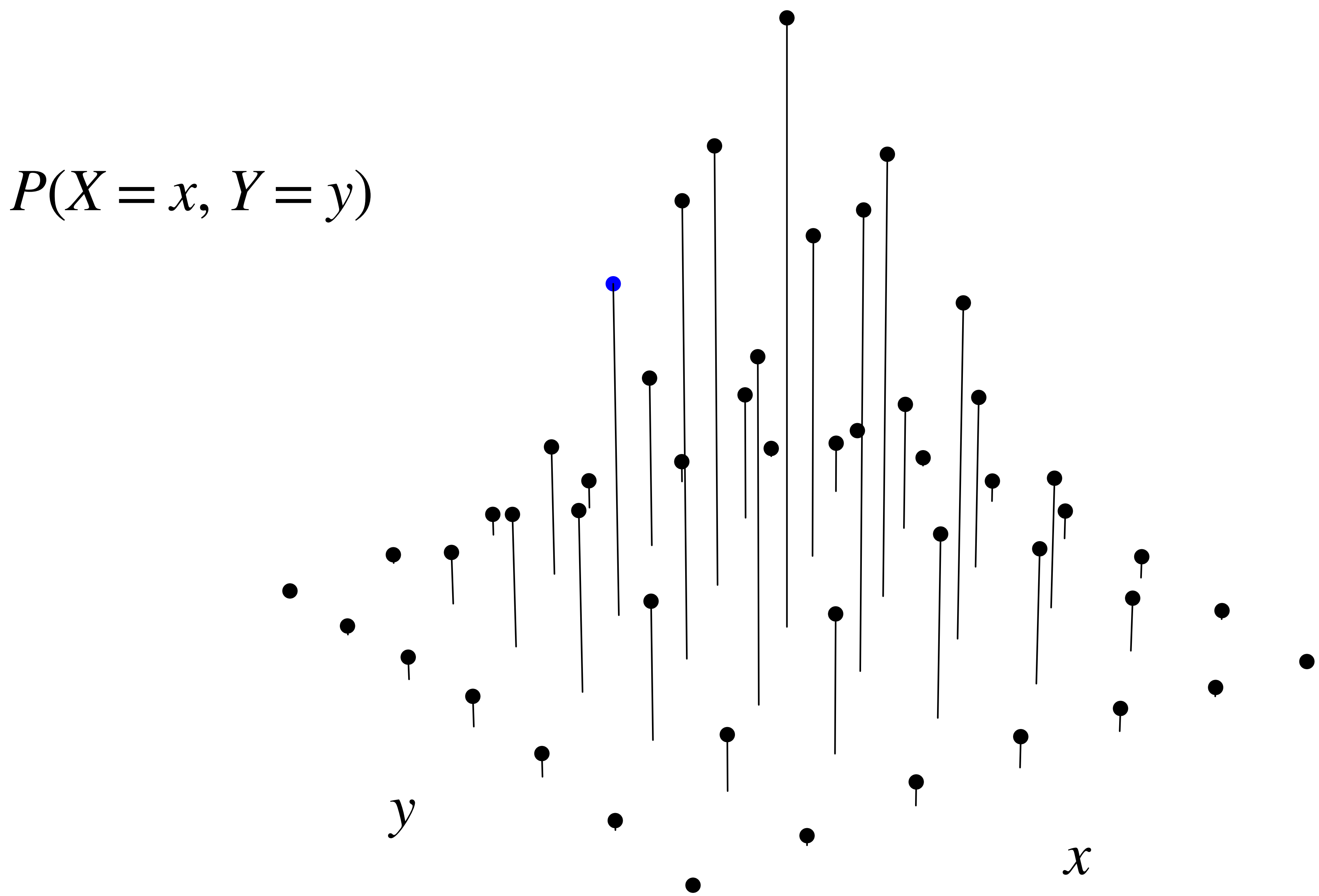
The joint CDF has jumps and flat regions, just like the univariate, so:
Definition 7.1.2 (Joint PMF). The **joint** PMF of discrete r.v.-s X and Y is the function $p_{X,Y}$ given by $p_{X,Y}(x, y) = P(X = x, Y = y)$.

Just as univariate PMFs, the joint should be nonnegative and sum to 1: $\Sigma_x \Sigma_y P(X = x, Y = y) = 1$

The probability of the event A is the sum over all points $(X, Y) \in A$:

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

Joint, marginal and conditional: discrete



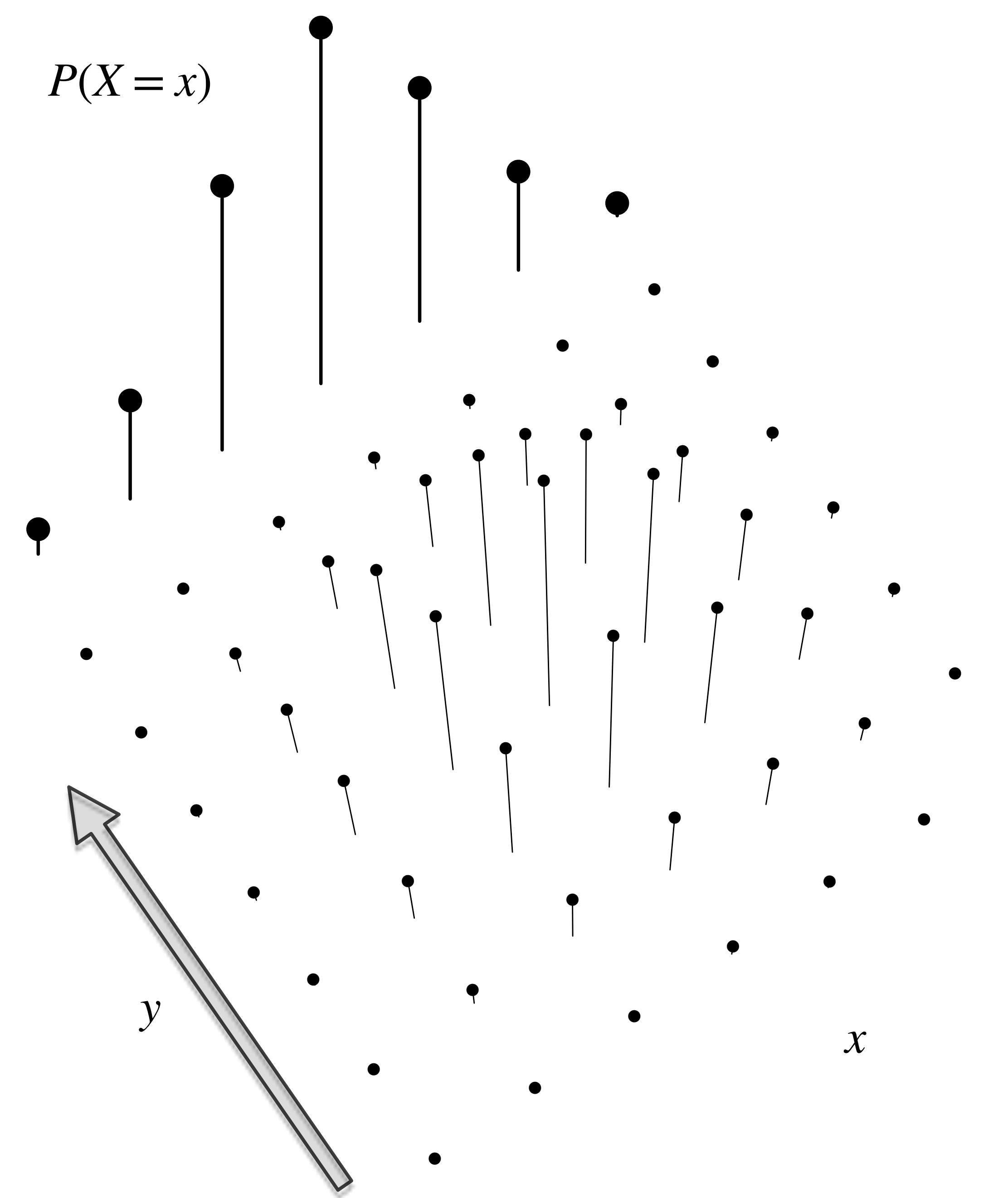
Here's an example of how a joint PMF of two discrete variables looks like.

Joint, marginal and conditional: discrete

Definition 7.1.3 (Marginal PMF). For discrete r.v.-s X and Y, the **marginal** PMF of X is $P(X=x) = \Sigma_y \, P(X=x, Y=y)$.

This gives the distribution of X individually, rather than jointly with Y.
Such summing over all possible values of Y is called **marginalising out Y.**

It is in general impossible to recover a joint distribution from two marginals!



Joint, marginal and conditional: discrete

Another way to go from joint to marginal is via the CDF:

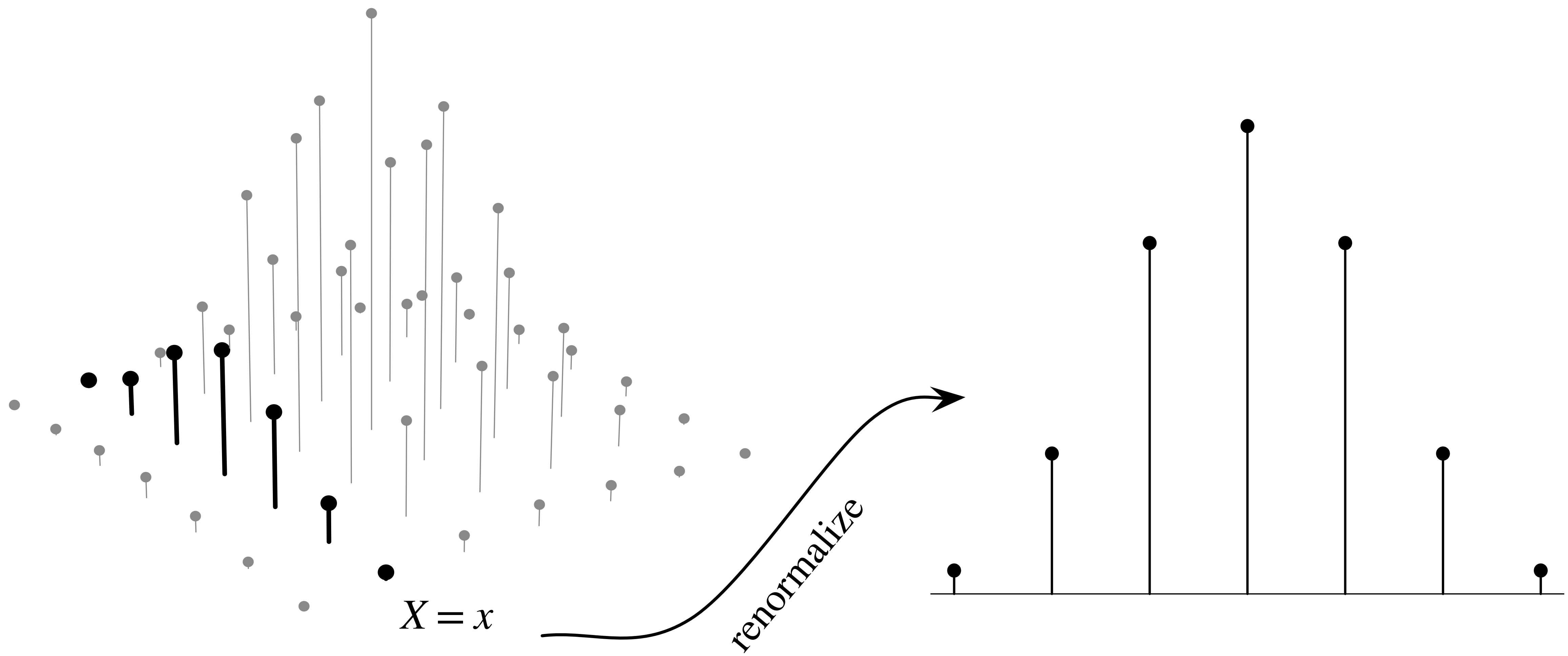
$$F_X(x) = P(X \leq x) = \lim_{y \to \infty} P(X \leq x, Y \leq y) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

Now suppose we observe the value of X and want to update our distribution of Y to reflect this:

Definition 7.1.4 (Conditional PMF). For discrete r.v.-s X and Y, the **conditional** PMF of Y given X is $P(Y = y \,|\, X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$

The conditional PMF is a valid PMF. So we can define, for example, the **conditional expectation** of Y given X: $E(Y \,|\, X = x)$.

Joint, marginal and conditional: discrete



Here's an example of how we obtain the conditional – take a "slice" of the joint PMF, and renormalise it (dividing by $P(X = x)$).

Joint, marginal and conditional: discrete

Definition 7.1.7 (Independence of discrete r.v.-s). R.v.-s X and Y are **independent** if $\forall x, y: \; F_{X,Y}(x, y) = F_X(x)F_Y(y)$. If X, Y are discrete, this is equivalent to $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x, y. Another form is $P(Y = y \,|\, X = x) = P(Y = y)$.

That is, for independent r.v.-s, the **joint** CDF (and PMF) is the product of the **marginal** CDFs (PMFs).
Or, equivalently, **conditional** PMFs are the same as **marginal** PMFs.

Joint, marginal and conditional:

Continuous case

Joint, marginal and conditional: continuous

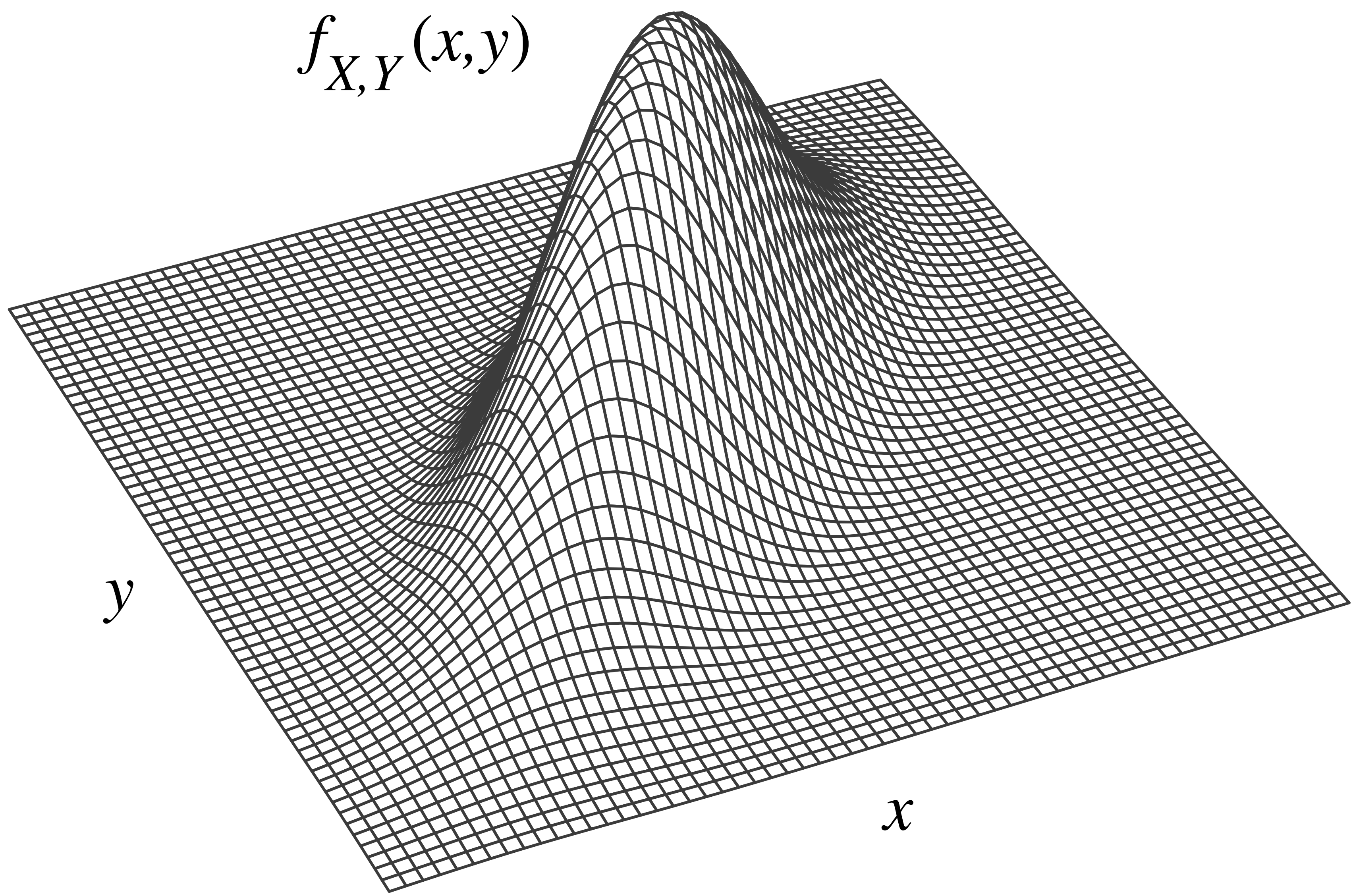
Formally, in order for X, Y to have a continuous joint distribution, we require the joint CDF $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ to be differentiable w.r.t. x and y. The partial derivative is:

Definition 7.1.13 (Joint PDF). If X, Y are continuous with joint CDF $F_{X,Y}$, their **joint** PDF is $f_{X,Y}(x, y) = \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

$f_{X,Y}$ should be nonnegative $f_{X,Y}(x, y) > 0$ and integrate to 1.

Probability of an event A – subset of the (X, Y) plane is the ***volume under*** the plot of the joint PDF.

Joint, marginal and conditional: continuous



Here’s an example of how a joint PDF of two continuous variables looks like.

Joint, marginal and conditional: continuous

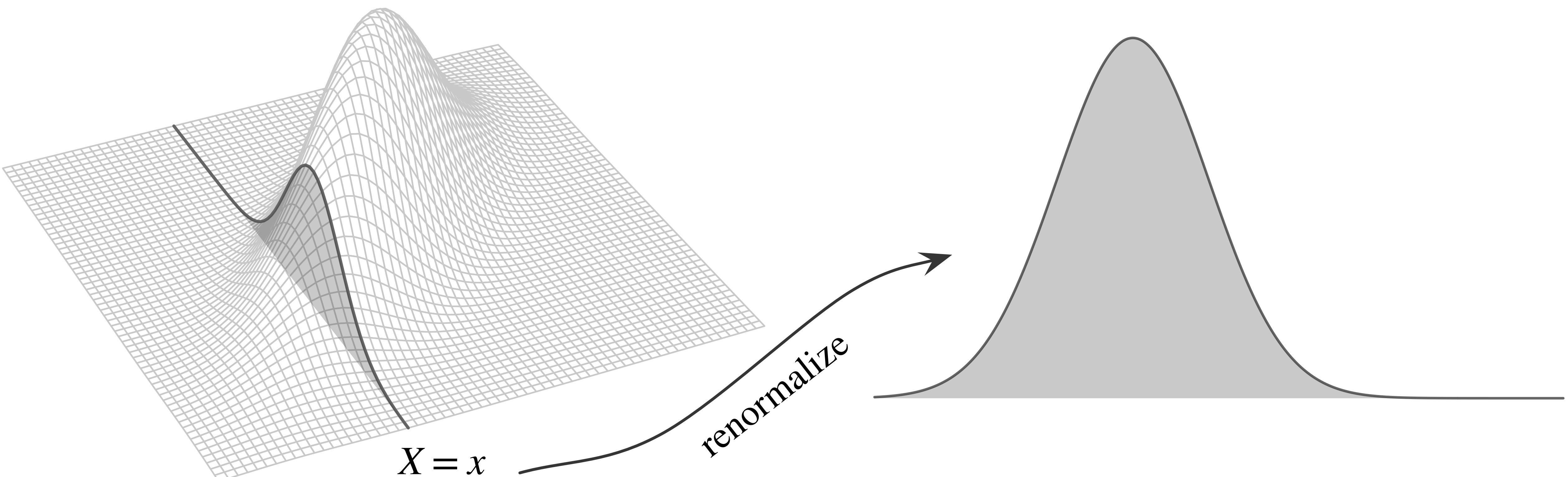
Definition 7.1.14 (Marginal PDF). For continuous r.v.-s X, Y with joint PDF $f_{X,Y}$ the **marginal** PDF of X is $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dy$

(So we marginalise Y out. This generalises to several r.v.s).

Definition 7.1.15 (Conditional PDF). For continuous r.v.s X, Y with joint PDF $f_{X,Y}$ the **conditional** PDF of Y given $X = x$ is

$f_{Y|X}(y\,|\,x) = \dfrac{f_{X,Y}(x, y)}{f_X(x)}$ for all x with $f_X(x) > 0$. To make $f_{Y|X}(y\,|\,x)$ well-defined for all real x, let $f_{Y|X}(y\,|\,x) = 0$ for such x that $f_X(x) = 0$.

Joint, marginal and conditional: continuous



Here's an example of how we obtain the conditional – take a "slice" of the joint PDF, and renormalise it (dividing by $f_X(x)$).

Joint, marginal and conditional: continuous

Theorem 7.1.18 (Continuous form of Bayes and LOTP). For continuous r.v.s X, Y, Bayes rule reads: $f_{Y|X}(y\,|\,x) = \dfrac{f_{X|Y}(x\,|\,y) f_Y(y)}{f_X(x)}$ for $f_X(x) > 0$. LOTP reads: $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f_{X|Y}(x\,|\,y) f_Y(y)\,dy$

Definition 7.1.19 (Independence of continuous r.v.s). R.v.s X, Y are **independent** if $\forall x, y$: $F_{X,Y}(x, y) = F_X(x)\,F_Y(y)$. In terms of the joint PDF, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. Another way is $f_{Y|X}(y\,|\,x) = f_Y(y)$.

Proposition 7.1.21 If the joint PDF of r.v.s X, Y, $f_{X,Y}$ factors as $f_{X,Y}(x, y) = g(x)\,h(y)$ – then X and Y are independent.

Joint, marginal and conditional: continuous

Example: Uniform on unit disk: $f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 = 1 \\ 0 & \text{otherwise} \end{cases}$

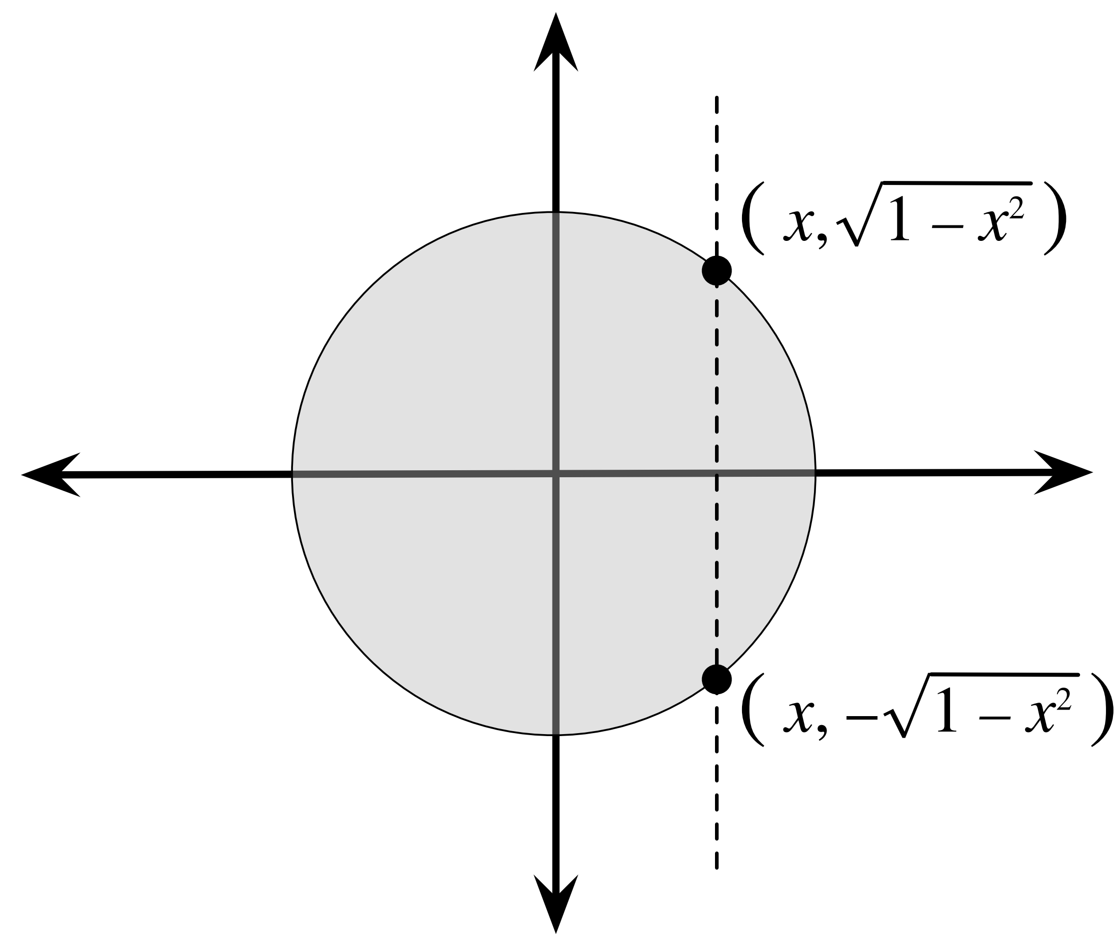
Note that X and Y are ***not*** independent.

The **marginal** is:

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\, dy = \frac{2}{\pi}\sqrt{1-x^2}$$

and the **conditional** is $f_{Y|X}(y\,|\,x) =$

$$= \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\frac{1}{\pi}}{\frac{2}{\pi}\sqrt{1-x^2}} = \frac{1}{2\sqrt{1-x^2}}$$



2D LOTUS

2D LOTUS

Theorem 7.2.1 (2D LOTUS). Let g be a function $g : \mathbb{R}^2 \to \mathbb{R}$. If X, Y are discrete, then $E(g(X, Y)) = \Sigma_x \Sigma_y g(x, y) P(X = x, Y = y)$

If continuous, then $E(g(X, Y)) = \displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y)\, dxdy$.

Example 7.2.2 (Expected distance b.w. two Uniforms). Let $X \sim$ Unif$(0,1)$, and same for Y. Find $E(|X - Y|)$.

Solution: $E(|X - Y|) = \displaystyle\int_0^1 \int_0^1 |x - y|\, dxdy =$

$$= \int_0^1 \int_y^1 (x - y)\, dxdy + \int_0^1 \int_0^y (y - x)\, dxdy = 2\int_0^1 \int_0^1 (x - y)\, dxdy = 1/3$$

Exercise: Find $E(|X - Y|)$ for $X, Y \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

Covariance and correlation

Covariance and correlation

Just like mean and variance provide single-number summaries of single r.v. distribution, covariance is a single-number summary of the joint distribution of two r.v.-s.

Definition 7.3.1 (Covariance). The **covariance** between X and Y is $\text{Cov}(X, Y) = E((X - EX)(Y - EY))$. By linearity, we get $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

Theorem 7.3.2 If X and Y are independent, they are uncorrelated.

Proof. Follows directly from $f_{XY}(x, y) = f_X(x)f_Y(y)$ for independent variables, so the double-integral of $E(XY)$ decomposes into a product of two integrals giving $E(X)\,E(Y)$.

Covariance and correlation

Here are some **properties** of covariance:

1) $\text{Cov}(X, X) = \text{Var}(X)$

2) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

3) $\text{Cov}(X, c) = 0$ for any constant c

4) $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$

5) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$

$\text{Cov}(X + Y, Z + W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$

6) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

Covariance and correlation

Correlation is just an appropriately normalised version of covariance:

Definition 7.3.4 (Correlation). The **correlation** between X and Y is

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var(X) Var(Y)}}}$$ (undefined if Var(X) or Var(Y) =0).

Theorem 7.3.5 (Correlation bounds). For any r.v.-s X and Y,

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Covariance and correlation