Lisa Yan and Jerry Cain
CS109

Lecture Notes #16
October 19, 2020

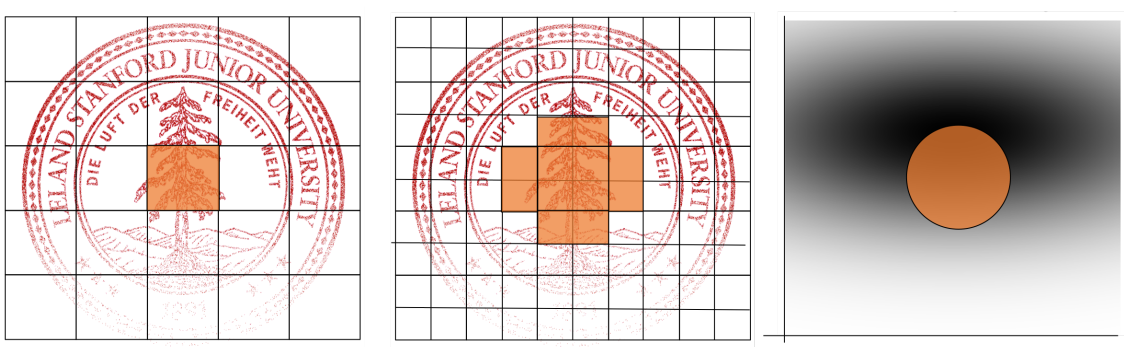
# Continuous Joint Distributions

Based on a chapter by Chris Piech and Lisa Yan

**Pre-recorded lecture**: Sections 1 to 3.
**In-lecture**: Section 5 (Exercises)
**Not covered**: Section 4 (Joint CDF)

## 1 Continuous Joint Distributions

Of course joint variables don't have to be discrete only, they can also be continuous. As an example: consider throwing darts at a dart board. Because a dart board is two dimensional, it is natural to think about the $X$ location of the dart and the $Y$ location of the dart as two random variables that are varying together (aka they are joint). However since x and y positions are continuous we are going to need new language to think about the likelihood of different places a dart could land. Just like in the non-joint case continuous is a little tricky because it isn't easy to think about the probability that a dart lands at a location defined to infinite precision. What is the probability that a dart lands at exactly ($X$=456.234231234122355, $Y$ = 532.12344123456)?:




Lets build some intuition by first starting with discretized grids. On the left of the image above you could imagine where your dart lands is one of 25 different cells in a grid. We could reason about the probabilities now! But we have lost all nuance about how likelihood is changing within a given cell. If we make our cells smaller and smaller we eventually will get a second derivative of probability: once again a probability density function. If we integrate under this joint-density function in both the x and y dimension we will get the probability that x takes on the values in the integrated range and y takes on the values in the integrated range!

Random variables $X$ and $Y$ are Jointly Continuous if there exists a Probability Density Function (PDF) $f_{X,Y}$ such that:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y)dy\ dx$$

Using the PDF we can compute marginal probability densities:

$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a,y)dy$$
$$f_Y(b) = \int_{-\infty}^{\infty} f_{X,Y}(x,b)dx$$

## 2 Independence with Multiple RVs (Continuous Case)

Two continuous random variables $X$ and $Y$ are called **independent** if:

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b) \text{ for all } a, b$$

This can be stated equivalently as:

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \text{ for all } a, b$$
$$f_{X,Y}(a,b) = f_X(a)f_Y(b) \text{ for all } a, b$$

More generally, if you can factor the joint density function, then your continuous random variables are independent:

$$f_{X,Y}(x,y) = h(x)g(y) \text{ where } -\infty < x, y < \infty$$

## 3 Bivariate Normal Distribution

Many times, we talk about multiple Normal (Gaussian) random variables, otherwise known as Multivariate Normal (Gaussian) distributions. Here, we talk about the two-dimensional case, called a Bivariate Normal Distribution. $X_1$ and $X_2$ follow a bivariate normal distribution if their joint PDF is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}.$$

We often write the distribution of the *vector* $\mathbf{X} = (X_1, X_2)$ as $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ is a *mean vector* and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ is a *covariance matrix*.

Note that $\rho$ is the correlation between $X_1$ and $X_2$, and $\sigma_1, \sigma_2 > 0$. We defer to Ross Chapter 6, Example 5d, for the full proof, but it can be shown that the marginal distributions of $X_1$ and $X_2$ are $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, respectively.

### *3.1 Example: Independent Normal RVs*

Let $\mathbf{X} = (X_1, X_2) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, a diagonal covariance matrix.

Noting that the correlation between $X_1$ and $X_2$ is $\rho = 0$:

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x_1-\mu_1)^2/(2\sigma_1^2)} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x_2-\mu_2)^2/(2\sigma_2^2)}$$

In other words, for Bivariate Normal RVs, if $\mathrm{Cov}(X_1, X_2) = 0$, then $X_1$ and $X_2$ are independent. Wild!

## 4 Joint CDFs

For two random variables $X$ and $Y$ that are jointly distributed, the joint cumulative distribution function $F_{X,Y}$ can be defined as

$$F_{X,Y}(a, b) = P(X \leq a, Y \leq b)$$

$$\begin{aligned}
F_{X,Y}(a, b) &= \sum_{x \leq a} \sum_{y \leq b} p_{X,Y}(x, y) && X, Y \text{ discrete} \\
F_{X,Y}(a, b) &= \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y) dy dx && X, Y \text{ continuous} \\
f_{X,Y}(a, b) &= \frac{\partial^2}{\partial a \partial b} F_{X,Y}(a, b) && X, Y \text{ continuous}
\end{aligned}$$

It can be shown via geometry that to calculate probabilities of joint distributions, we can use the CDF as follows, for both jointly discrete and jointly continuous RVs:

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1)$$


In image processing, a Gaussian blur is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise.

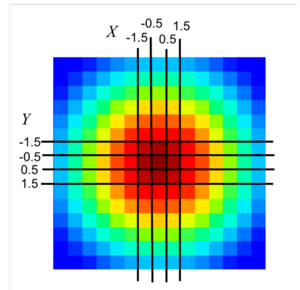
Gaussian blurring with StDev = 3, is based on a joint probability distribution:

**Joint PDF**

$$f_{X,Y}(x,y) = \frac{1}{2\pi \cdot 3^2} e^{-\frac{x^2+y^2}{2\cdot 3^2}}$$

**Joint CDF**

$$F_{X,Y}(x,y) = \Phi\left(\frac{x}{3}\right) \cdot \Phi\left(\frac{y}{3}\right)$$




Used to generate this weight matrix

## 4.1 *Example: Gaussian Blur*

Lets make a weight matrix used for Gaussian blur. In the weight matrix, each location in the weight matrix will be given a weight based on the probability density of the area covered by that grid square in a Bivariate Normal of independent $X$ and $Y$, each zero mean with variance $\sigma^2$. For this example lets blur using $\sigma = 3$.

Each pixel is given a weight equal to the probability that X and Y are both within the pixel bounds. The center pixel covers the area where $-0.5 \leq x \leq 0.5$ and $-0.5 \leq y \leq 0.5$ What is the weight of the center pixel?

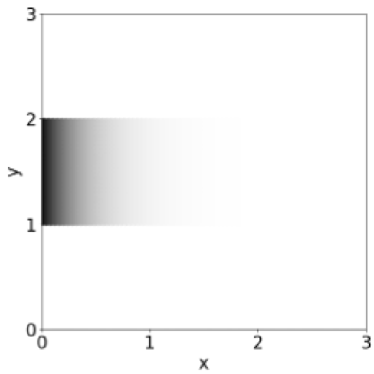
$$\begin{aligned}
&P(-0.5 < X < 0.5, -0.5 < Y < 0.5) \\
=&P(X < 0.5, Y < 0.5) - P(X < 0.5, Y < -0.5) \\
&- P(X < -0.5, Y < 0.5) + P(X < -0.5, Y < -0.5) \\
=&\phi\left(\frac{0.5}{3}\right) \cdot \phi\left(\frac{0.5}{3}\right) - 2\phi\left(\frac{0.5}{3}\right) \cdot \phi\left(\frac{-0.5}{3}\right) \\
&+ \phi\left(\frac{-0.5}{3}\right) \cdot \phi\left(\frac{-0.5}{3}\right) \\
=&0.5662^2 - 2 \cdot 0.5662 \cdot 0.4338 + 0.4338^2 = 0.206
\end{aligned}$$

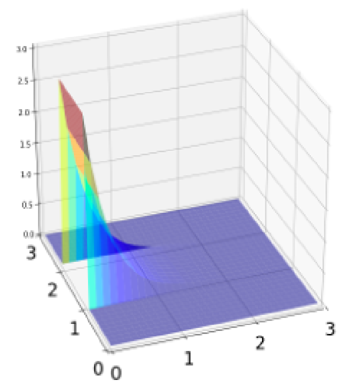
# 5 Exercises

## 5.1 *Example: Jointly continuous random variables*

Suppose that $X$ and $Y$ have the joint PDF $f_{X,Y}(x,y) = 3e^{-3x}$ where $0 < x < \infty$ and $1 < y < 2$. $f_{X,Y}(x,y) = 0$ outside of this support. We graph the joint PDF below:

**Independence:** $X$ and $Y$ are independent because the joint PDF can be separated into $g(x) = 3Ce^{-3x}$ and $h(y) = 1/C$, where $C$ is a constant and $(x,y)$ are in the support. We can also intuitively look at the joint PDF and note that for all values of $y$, $x$ has the same type of exponential curve slope; vice versa, for all values of $x$, $y$ looks to have even weight in the range $1 < y < 2$.



**Marginal distributions:** We can compute the marginal pdf of $Y$ by observing that the only value of $C$ for which $h(y)$ is a valid PDF is $C = 1$. Therefore $Y \sim \text{Uni}(a = 1, b = 2)$, and similarly $X \sim \text{Exp}(\lambda = 3)$.

**Expectation of sum:** Suppose we wanted to compute $E[X + Y]$. We could compute this using LOTUS as $E[X + Y] = \int_1^2 \int_0^\infty (x + y)3e^{-3x}$. However, it's probably easier to use the marginal distributions of $X$ and $Y$ (since we know them) and compute using linearity of expectation as $E[X + Y] = E[X] + E[Y] = 1/3 + 3/2$.

## 5.2 *Example: The joy of meetings*

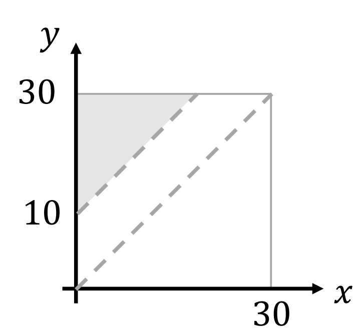
Two people set up a meeting time. Each person arrives independently at a time uniformly between 12pm and 12:30pm. What is the probability that the first person to arrive waits more than 10 minutes for the other person?

**Solution:** Let $X$ and $Y$ be the numbers of minutes past 12pm that person 1 and person 2 arrive, respectively. $X \sim \text{Uni}(0, 30)$ and $Y \sim \text{Uni}(0, 30)$.

We would like to compute the probability of two mutually exclusive events: either person 1 arrives first and waits more than 10 minutes, in which case $X + 10 < Y$, or person 4 arrives first and waits more than 10 minutes, in which case $Y + 10 < X$. By symmetry, $P(X + 10 < Y) = P(Y + 10 < X)$ (note that this phrase "by symmetry" means that the joint PDF $f_{X,Y}(x, y) = f_{Y,X}(y, x)$), and therefore we would like to compute

$$2P(X + 10 < Y) = 2 \cdot \iint_{x+10<y} f_{X,Y}(x, y)\, dxdy = 2 \cdot \iint_{\substack{x+10<y \\ 0 \le x, y, \le 30}} \left(\frac{1}{30}\right)^2 dxdy$$

The figure on the right shows the region over which we integrate—our outer integral bound is $y \in [10, 30]$ and our inner integral bound is $x \in [0, y-10]$. We can also determine this mathematically, but it's more complicated: The lower bound on $y$ arises because $y > x + 10$ for all $x$ in our target integral, and our minimum $x$ is 0. The bounds on $x$ arise from treating the $y$ from the outer integral as a constant.

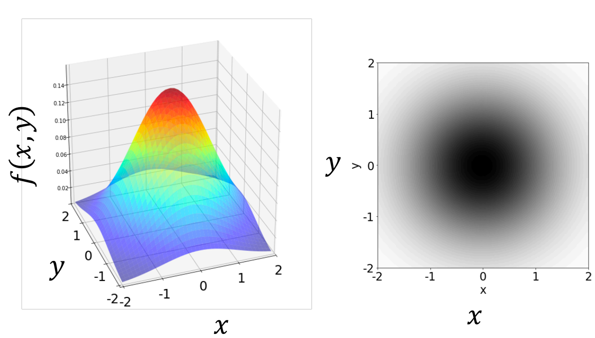


Our bounds on $x$ must respect both the bounds $0 \leq x \leq 30$ and $x < y - 10$, and therefore we choose $x \in [0, y - 10]$ because $y - 10 \leq 30$ for all $y$ in the PDF's support.

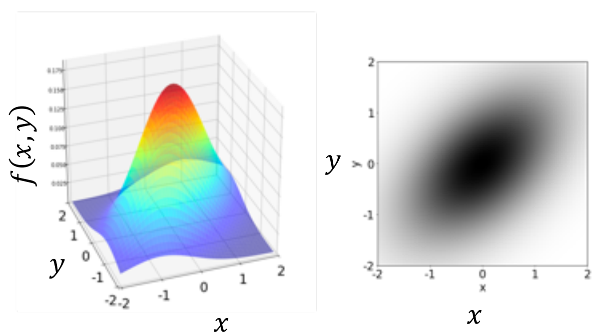
$$2P(X + 10 < Y) = \frac{2}{30^2} \int_{10}^{30} \int_{0}^{y-10} dxdy = \frac{2}{30^2} = \int_{10}^{30} (y - 10)dy = \frac{4}{9}$$

### 5.3 Example: Visualizing the Bivariate Normal Distribution

Suppose that $X$ and $Y$ are distributed as a bivariate normal where $E[X] = E[Y] = 0$. The below figure shows four possible joint PDFs of $(X, Y)$ with different covariance matrices $\Sigma$. We include both the 3-D view and top-down view for each joint PDF.

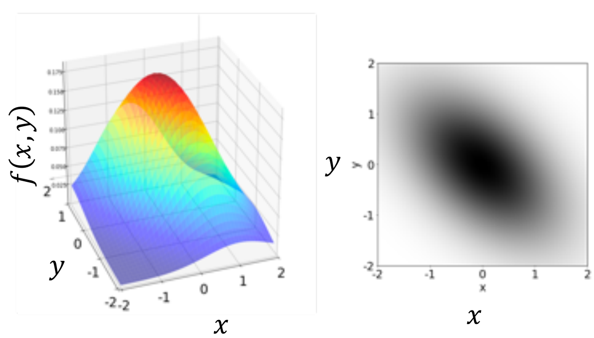



(a) $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

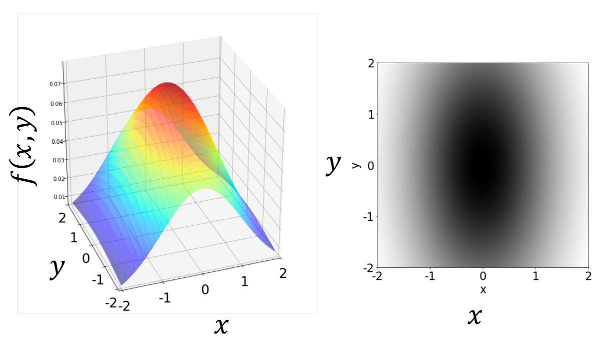



(b) $\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

$x$




(c) $\Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$




(d) $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

### 5.4 Example: Integral practice

Let $X$ and $Y$ be two continuous random variables with joint PDF:

$$f(x, y) = \begin{cases} 4xy & 0 \leq x, y, \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X \leq Y)$?

**Solution:** The big challenge is to determine the bounds of the double integral. If we are looking for the region of the support where $X \le Y$, then we could for all values of $y \in [0, 1]$, we integrate $x$ from 0 to $y$ (thus guaranteeing that $x \le y$. Alternatively, for all values of $x \in [0, 1]$, we integrate $y$ from $x$ to 1. We take the former approach below.

$$\begin{aligned} P(X \le Y) &= \iint_{\substack{x \le y, \\ 0 \le x, y, \le 1}} 4xy\, dxdy = \int_{y=0}^{1} \int_{x \le y} 4xy\, dxdy = \int_{y=0}^{1} \int_{x=0}^{y} 4xy\, dxdy \\ &= \int_{y=0}^{1} 4y \left[\frac{x^2}{2}\right]_0^y dy = \int_{y=0}^{1} 2y^3 dy = \left[\frac{2}{4} y^4\right]_0^1 = \frac{1}{2} \end{aligned}$$

We could also have noticed that $P(X \le Y) + P(X > Y) = 1$ and because the joint PDF is symmetric, $P(X \le Y) = P(Y \le X) = P(Y < X)$. Therefore $2P(X \le Y) = 1$, and thus $P(X \le Y) = 1/2$.

## *5.5 Example: Imperfection on Disk*

Suppose that you have a disk surface, modeled as a circle of radius $R$. Suppose that you know that there is a single point imperfection uniformly distributed on the disk. Therefore the coordinates $(X, Y)$ of this imperfection is distributed according to the following joint PDF:

$$f(x, y) = \begin{cases} \frac{1}{\pi R^2} & x^2 + y^2 \le R^2 \\ 0 & \text{otherwise} \end{cases}$$

What are the marginal distributions of $X$ and $Y$? Are $X$ and $Y$ independent?

**Solution:** To compute $f_X(x)$, the marginal PDF of $X$, we note that we must integrate over $y$ in the support, where $x^2 + y^2 \le R^2$ and therefore $y \in [-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}]$ for $-R \le x \le R$.

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

By symmetry, we observe that we could switch $x$ and $y$ above and obtain the marginal PDF of $y$:

$$f_Y(y) = \frac{2\sqrt{R^2 - y^2}}{\pi R^2} \text{where } -R \le y \le R$$

$X$ and $Y$ are not independent; they are dependent because $f_{X,Y}(x, y) \ne f_X(x) f_Y(y)$.