Lisa Yan
CS109

Lecture Notes #17
May 13, 2020

Continuous Joint Distributions, Part II

Based on a chapter by Chris Piech and Lisa Yan

Convolution: Sum of independent random variables

Remember how deriving the sum of two independent Poisson random variables was tricky? When we move into integral land, the concept of convolution still carries over, and once you get a handle on notation, then computing the sum of two independent, jointly continuous random variables becomes fun. For some definition of fun. . .

Independent Normals

Let's start with one common case that has a nice form but a difficult derivation:

For any two normal random variables $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ the sum of those two random variables is another normal: $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

We won't derive the approach here, but it involves exponents, integrals, and completing the square (algebra throwback!).

General Independent Case

For two general independent random variables (aka cases of independent random variables that don't fit the above special situations) you can calculate the CDF or the PDF of the sum of two random variables using the following convolution formulas:

$$F_{X+Y}(a) = P(X + Y \leq a) = \int_{y=-\infty}^{\infty} F_X(a - y) f_Y(y) dy$$

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f_X(a - y) f_Y(y) dy$$

These is a direct analogy to the discrete case where you replace the integrals with sums and change notation for CDF and PDF.

Example 1

What is the PDF of $X + Y$ for independent uniform random variables $X \sim \text{Uni}(0, 1)$ and $Y \sim \text{Uni}(0, 1)$? First plug in the equation for general convolution of independent random variables:

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y) f_Y(y) dy$$

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y) dy \qquad \text{because } f_Y(y) = 1$$

It turns out that is not the easiest thing to integrate. By trying a few different values of a in the range $[0, 2]$ we can observe that the PDF we are trying to calculate is discontinuous at the point $a = 1$ and thus will be easier to think about as two cases: $a < 1$ and $a > 1$. If we calculate f_{X+Y} for both cases and correctly constrain the bounds of the integral we get simple closed forms for each case:

$$f_{X+Y}(a) = \begin{cases} a & \text{if } 0 < a \leq 1 \\ 2 - a & \text{if } 1 < a \leq 2 \\ 0 & \text{else} \end{cases}$$

Conditional Distributions (Continuous case)

The conditional probability density function might look a bit wonky, but it works!

Continuous

The conditional probability density function (PDF) for the continuous case:

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

The conditional cumulative density function (CDF) for the continuous case:

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

At first glance, the conditional density function seems to violate any notion of stoichiometric units of probability. Let us verify this with our understanding of discrete probability. Recall that for tiny epsilon ε, we can approximate $P(|X - x| \leq \frac{\varepsilon}{2}) = P(x - \frac{\varepsilon}{2} \leq X \leq x + \frac{\varepsilon}{2}) = \int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} f_X(a) da \approx f_X(x)\varepsilon$. This extends to the joint variable case: $P(|X - x| \leq \frac{\varepsilon_X}{2}, |Y - y| \leq \frac{\varepsilon_y}{2}) \approx f_{X,Y}(x, y)\varepsilon_X \varepsilon_Y$.

$$\begin{aligned} P\left(|X - x| \leq \frac{\varepsilon_X}{2} \middle| |Y - y| \leq \frac{\varepsilon_Y}{2}\right) &= \frac{P(|X - x| \leq \frac{\varepsilon_X}{2}, |Y - y| \leq \frac{\varepsilon_Y}{2})}{P(|Y - y| \leq \frac{\varepsilon_Y}{2})} && \text{(def. cond. prob.)} \\ &\approx \frac{f_{X,Y}(x, y)\varepsilon_X \varepsilon_Y}{f_Y(y)\varepsilon_Y} = \frac{f_{X,Y}(x, y)}{f_Y(y)}\varepsilon_X = f_{X|Y}(x|y)\varepsilon_X \end{aligned}$$

Conditional expectation (Continuous case)

Conditional expectation in the continuous case is a direct analogy to what we saw in the discrete case:

Let X and Y be jointly continuous random variables. We define the conditional expectation of X given $Y = y$ to be:

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

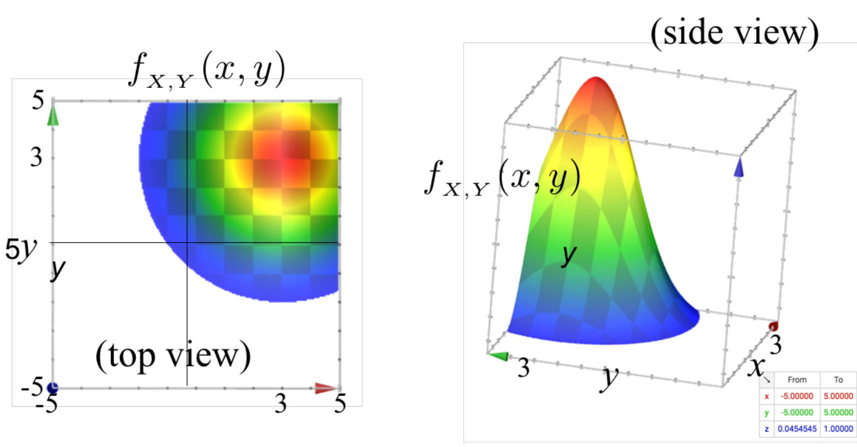
Example 1

In this example we are going to explore the problem of tracking an object in 2D space. The object exists at some (x, y) location, however we are not sure exactly where! Thus we are going to use random variables X and Y to represent location.

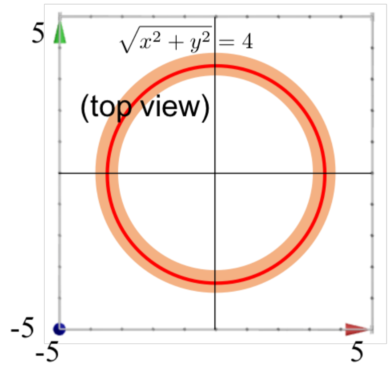
We have a prior belief about where the object is. In this example our prior both X and Y as normals which are independently distributed with mean 3 and variance 4. First let's write the prior belief as a joint probability density function

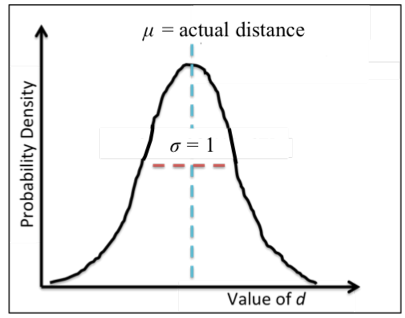
$$\begin{aligned} f(X = x, Y = y) &= f(X = x) \cdot f(Y = y) && \text{In the prior X and Y are independent} \\ &= \frac{1}{\sqrt{2 \cdot 4 \cdot \pi}} \cdot e^{-\frac{(x-3)^2}{2 \cdot 4}} \cdot \frac{1}{\sqrt{2 \cdot 4 \cdot \pi}} \cdot e^{-\frac{(y-3)^2}{2 \cdot 4}} && \text{Using the PDF equation for normals} \\ &= K_1 \cdot e^{-\frac{(x-3)^2+(y-3)^2}{8}} && \text{All constants are put into } K_1 \end{aligned}$$

This combinations of normals is called a bivariate distribution. Here is a visualization of the PDF of our prior.



The interesting part about tracking an object is the process of updating your belief about it's location based on an observation. Let's say that we get an instrument reading from a sonar that is sitting on the origin. The instrument reports that the object is 4 units away. Our instrument is not perfect: if the true distance was t units away, than the instrument will give a reading which is normally distributed with mean t and variance 1. Let's visualize the observation:





Based on this information about the noisiness of our prior, we can compute the conditional probability of seeing a particular distance reading D, given the true location of the object X, Y. If we knew the object was at location (x, y), we could calculate the true distance to the origin $\sqrt{x^2 + y^2}$ which would give us the mean for the instrument Gaussian:

$$
\begin{aligned}
f(D = d|X = x, Y = y) &= \frac{1}{\sqrt{2 \cdot 1 \cdot \pi}} \cdot e^{-\frac{\left(d-\sqrt{x^2+y^2}\right)^2}{2\cdot 1}} && \text{Normal PDF where } \mu = \sqrt{x^2 + y^2} \\
&= K_2 \cdot e^{-\frac{\left(d-\sqrt{x^2+y^2}\right)^2}{2\cdot 1}} && \text{All constants are put into } K_2
\end{aligned}
$$

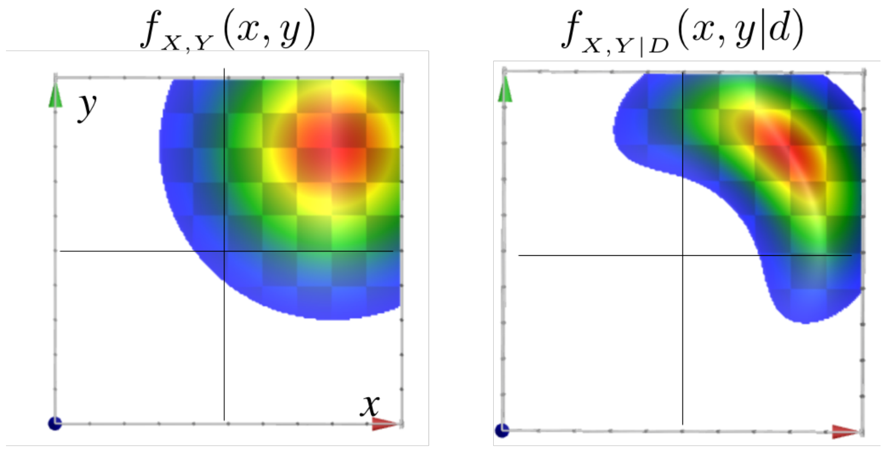
How about we try this out on actual numbers. How much more likely is an instrument reading of 1 compared to 2, given that the location of the object is at (1, 1)?

$$
\begin{aligned}
\frac{f(D = 1|X = 1, Y = 1)}{f(D = 2|X = 1, Y = 1)} &= \frac{K_2 \cdot e^{-\frac{\left(1-\sqrt{1^2+1^2}\right)^2}{2\cdot 1}}}{K_2 \cdot e^{-\frac{\left(2-\sqrt{1^2+1^2}\right)^2}{2\cdot 1}}} && \text{Substituting into the conditional PDF of D} \\
&= \frac{e^0}{e^{-1/2}} \approx 1.65 && \text{Notice how the } K_2 \text{ cancel out}
\end{aligned}
$$

At this point we have a prior belief and we have an observation. We would like to compute an updated belief, given that observation. This is a classic Bayes' formula scenario. We are using joint continuous variables, but that doesn't change the math much, it just means we will be dealing with densities instead of probabilities:

$$
\begin{aligned}
f(X = x, Y = y|D = 4) &= \frac{f(D = 4|X = x, Y = y) \cdot f(X = x, Y = y)}{f(D = 4)} && \text{Bayes using densities} \\
&= \frac{K_1 \cdot e^{-\frac{[4-\sqrt{x^2+y^2})^2]}{2}} \cdot K_2 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}}{f(D = 4)} && \text{Substituting for prior and update} \\
&= \frac{K_1 \cdot K_2}{f(D = 4)} \cdot e^{-\left[\frac{[4-\sqrt{x^2+y^2})^2]}{2} + \frac{[(x-3)^2+(y-3)^2]}{8}\right]} && f(D = 4) \text{ is a constant w.r.t. } (x, y) \\
&= K_3 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{[(x-3)^2+(y-3)^2]}{8}\right]} && K_3 \text{ is a new constant}
\end{aligned}
$$

Wow! That looks like a pretty interesting function! You have successfully computed the updated belief. Let's see what it looks like. Here is a figure with our prior on the left and the posterior on the right: How beautiful is that! Its like a 2D normal distribution merged with a circle. But wait, what



Prior Posterior

about that constant! We do not know the value of K_3 and that is not a problem for two reasons: the first reason is that if we ever want to calculate a relative probability of two locations, K_3 will cancel out. The second reason is that if we really wanted to know what K_3 was, we could solve for it.

This math is used every day in millions of applications. If there are multiple observations the equations can get truly complex (even worse than this one). To represent these complex functions often use an algorithm called particle filtering.

Example 2

Let's say we have two independent random Poisson variables for requests received at a web server in a day: X = # requests from humans/day, $X \sim Poi(\lambda_1)$ and Y = # requests from bots/day, $Y \sim Poi(\lambda_2)$. Since the convolution of Poisson random variables is also a Poisson we know that the total number of requests $(X + Y)$ is also a Poisson $(X + Y) \sim Poi(\lambda_1 + \lambda_2)$. What is the probability of having k human requests on a particular day given that there were n total requests?

$$
\begin{aligned}
P(X = k|X + Y = n) &= \frac{P(X = k, Y = n - k)}{P(X + Y = n)} = \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} \\
&= \frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \cdot \frac{n!}{e^{1(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \\
&\sim Bin\left(n, \frac{\lambda_2}{\lambda_1 + \lambda_2}\right)
\end{aligned}
$$