Chris Piech
CS109

Lecture #14
May 3rd, 2017

Convolution and Conditionals with Variables

Convolution of Distributions

Convolution is the result of adding two different random variables together. For some particular random variables computing convolution has intuitive closed form equations. Importantly convolution is the sum of the random variables themselves, not the addition of the probability density functions (PDF)s that correspond to the random variables.

Independent Binomials with equal p

For any two Binomial random variables with the same "success" probability: $X \sim Bin(n_1, p)$ and $Y \sim Bin(n_2, p)$ the sum of those two random variables is another binomial: $X + Y \sim Bin(n_1 + n_2, p)$. This does not hold when the two distribution have different parameters p.

Independent Poissons

For any two Poisson random variables: $X \sim Poi(\lambda_1)$ and $Y \sim Poi(\lambda_2)$ the sum of those two random variables is another Poisson: $X + Y \sim Poi(\lambda_1 + \lambda_2)$. This holds when λ_1 is not the same as λ_2.

Independent Normals

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: $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

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 formulas:

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

There are direct analogies in the discrete case where you replace the integrals with sums and change notation for CDF and PDF.

Example 1

Calculate the PDF of $X + Y$ for independent uniform random variables $X \sim Uni(0,1)$ and $Y \sim Uni(0,1)$? 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

Before we looked at conditional probabilities for events. Here we formally go over conditional probabilities for random variables. The equations for both the discrete and continuous case are intuitive extensions of our understanding of conditional probability:

Discrete

The conditional probability mass function (PMF) for the discrete case:

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

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

$$F_{X|Y}(a|y) = P(X \leq a|Y = y) = \frac{\sum_{x \leq a} p_{X,Y}(x,y)}{p_Y(y)} = \sum_{x \leq a} p_{X|Y}(x|y)$$

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:

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

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?

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