Lisa Yan and Jerry Cain
CS109

Lecture Notes #12
October 9, 2020

Independent Random Variables

Based on a chapter by Chris Piech

Pre-recorded lecture: Sections 1 and 3.
In-lecture: Section 2 and exercises.

1 Independence with Multiple RVs (Discrete Case)

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

$$P(X = x, Y = y) = P(X = x)P(Y = y) \text{ for all } x, y$$

Intuitively: knowing the value of X tells us nothing about the distribution of Y. If two variables are not independent, they are called dependent. This is a similar conceptually to independent events, but we are dealing with multiple *variables*. Make sure to keep your events and variables distinct.

2 Symmetry of Independence

Independence is symmetric. That means that if random variables X and Y are independent, X is independent of Y and Y is independent of X. This claim may seem meaningless but it can be very useful. Imagine a sequence of events $X_1, X_2, \ldots$. Let A_i be the event that X_i is a "record value" (eg it is larger than all previous values). Is A_{n+1} independent of A_n? It is easier to answer that A_n is independent of A_{n+1}. By symmetry of independence both claims must be true.

3 Sums of Independent Random Variables

Independent Binomials with equal p

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

Independent Poissons

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

3.1 Example: Web requests

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?

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

3.2 Example: Web requests, redux

Let N be the number of requests to a web server/day and that $N \sim \text{Poi}(\lambda)$. Each request comes from a human (probability = p) or from a "bot" (probability = $(1 - p)$), independently. Define X to be the number of requests from humans/day and Y to be the number of requests from bots/day.

Since requests come in independently, the probability of X conditioned on knowing the number of requests is a Binomial. Specifically, conditioned:

$$\begin{aligned}
(X|N) &\sim \text{Bin}(N, p) \\
(Y|N) &\sim \text{Bin}(N, 1 - p)
\end{aligned}$$

Calculate the probability of getting exactly i human requests and j bot requests. Start by expanding using the chain rule:

$$P(X = i, Y = j) = P(X = i, Y = j|X + Y = i + j)P(X + Y = i + j)$$

We can calculate each term in this expression:

$$\begin{aligned}
P(X = i, Y = j|X + Y = i + j) &= \binom{i+j}{i} p^i (1 - p)^j \\
P(X + Y = i + j) &= e^{-\lambda}\frac{\lambda^{i+j}}{(i + j)!}
\end{aligned}$$

Now we can put those together and simplify:

$$P(X = i, Y = j) = \binom{i+j}{i} p^i (1 - p)^j e^{-\lambda}\frac{\lambda^{i+j}}{(i + j)!}$$

As an exercise you can simplify this expression into two independent Poisson distributions.

Convolution: Sum of independent random variables

So far, we have had it easy: If our two independent random variables are both Poisson, or both Binomial with the same probability of success, then their sum has a nice, closed form. In the general case, however, the distribution of two independent random variables can be calculated as a **convolution** of probability distributions.

For two independent random variables, you can calculate the CDF or the PDF of the sum of two random variables using the following formulas:

$$F_{X+Y}(n) = P(X + Y \leq n) = \sum_{k=-\infty}^{\infty} F_X(k) F_Y(n-k)$$

$$p_{X+Y}(n) = \sum_{k=-\infty}^{\infty} p_X(k) p_Y(n-k)$$

Most importantly, convolution is the process of finding the sum of the random variables themselves, and not the process of adding together probabilities.

Example: Proof of sum of Poissons

Let's go about proving that the sum of two independent Poisson random variables is also Poisson. Let $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ be two independent random variables, and $Z = X + Y$. What is $P(Z = n)$?

$$\begin{aligned}
P(Z = n) = P(X + Y = n) &= \sum_{k=-\infty}^{\infty} P(X = k) P(Y = n-k) && \text{(Convolution)} \\
&= \sum_{k=0}^{n} P(X = k) P(Y = n-k) && \text{(Range of } X \text{ and } Y\text{)} \\
&= \sum_{k=0}^{n} e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} && \text{(Poisson PMF)} \\
&= e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n && \text{(Binomial theorem)}
\end{aligned}$$

Note that the Binomial Theorem (which we did not cover in this class, but is often used in contexts like expanding polynomials) says that for two numbers a and b and positive integer n, $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.