Topic 15: Multiple Random Variables

02-680: Essentials of Mathematics and Statistics

December 3, 2024

Sometimes we need to deal with the interactions of more than one random variable.

Consider tossing a (fair) coin 3 times, and define two random variables:

- X — the number of heads in the first toss
- Y — the number of heads in all 3 tosses

We want to know the joint probability (that is, the probability $p(X = x, Y = y)$):

		Y			
		0	1	2	3
X	0	1/8	1/4	1/8	0
	1	0	1/8	1/4	1/8

1 Marginal Probabilities

What if we're given the joint probability of some events and want to find out the underlying probability of the events.

Example. Assume we have two ***unfair*** coins, which we will model as random variables X and Y,

and we're given the following joint probabilities:

		Y		
		`Heads`	`Tails`	
X	`Heads`	1/10	2/10	3/10
	`Tails`	3/10	4/10	7/10
		4/10	6/10	

If we want to determine, say $P(X = \texttt{Tails})$, it turns out we can some over all the possibilities of Y:

$$P(X = \texttt{Tails}) = \sum_{y \in \{\texttt{Heads},\texttt{Tails}\}} p(X = \texttt{Tails}, Y = y) = \frac{3}{10} + \frac{4}{10} = \frac{7}{10}$$

Doing all the math, both coins are biased toward `Tails`, the first coin (X) more-so.

1.1 Continuous Random Variables

For continuous random variables, this again turns from sums into integrals. Assume you're given a joint distribution function $f(x, y)$. We can, as before, find the joint probability for a range as a (double) integral:

$$p(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx$$

And to get a marginal distribution function

$$f(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Example. Lets say we're drawing points uniformly over the unit square, in this case

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

So if we wan to find

$$p\left(X \leq \frac{1}{2}, Y \leq \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} 1 dy dx = \left((xy)\big|_0^{\frac{1}{2}}\right)\Big|_0^{\frac{1}{2}} = \left(x \cdot \frac{1}{2} - x \cdot 0\right)\Big|_0^{\frac{1}{2}}$$
$$= \left(\frac{1}{2} \cdot \frac{1}{2} - \frac{1}{2} \cdot 0\right) - \left(0 \cdot \frac{1}{2} - 0 \cdot 0\right) = \frac{1}{4}$$

2 Independence of Random Variables

Similar to independence for event probabilities, two random variables X and Y are independent iff

$$\forall x, y : \quad p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$$

Work left to the reader: one of the situations below describes independent random variables the other does not.

		Y: Heads	Y: Tails	
X	Heads	1/4	1/4	1/2
	Tails	1/4	1/4	1/2
		1/2	1/2	

		Y: Heads	Y: Tails	
X	Heads	1/2	0	1/2
	Tails	0	1/2	1/2
		1/2	1/2	

3 Conditional Distributions

Similar to independence for event probabilities, for two random variables X and Y :

$$p(X \mid Y) = \frac{p(X, Y)}{p(Y)}$$

BRCA1 mutation G_1	*P53* mutation G_2	develops cancer C	$p(G_1, G_2, C)$
N	N	N	0.24
N	N	Y	0.01
N	Y	N	0.20
N	Y	Y	0.05
Y	N	N	0.15
Y	N	Y	0.10
Y	Y	N	0.05
Y	Y	Y	0.20

Compute $p(C \mid G_1, G_2)$.

$$p(C \mid G_1, G_2) = \frac{p(G_1, G_2, C)}{p(G_1, G_2)} = \frac{p(G_1, G_2, C)}{p(G_1, G_2, C = \texttt{N}) + p(G_1, G_2, C = \texttt{Y})}$$

Note that

$$p(G_1 = a, G_2 = b) = \frac{\sum_{c \in \{\texttt{Y},\texttt{N}\}} p(G_1 = a, G_2 = b, C = c)}{\sum_{g_1 \in \{\texttt{Y},\texttt{N}\}} \sum_{g_2 \in \{\texttt{Y},\texttt{N}\}} \sum_{c \in \{\texttt{Y},\texttt{N}\}} p(G_1 = g_1, G_2 = g_2, C = c)}$$

but because the denominator is 1 it is left out of the equation above for ease exposition.

G_1	G_2	$p(C = 1 \mid G_1, G_2)$
N	N	$\frac{0.01}{0.24+0.01} = 0.04$
N	Y	$\frac{0.05}{0.20+0.05} = 0.20$
Y	N	$\frac{0.10}{0.15+0.10} = 0.40$
Y	Y	$\frac{0.20}{0.05+0.20} = 0.80$

Useful References

Wasserman. “All of Statistics: A Concise Course in Statistical Inference” §§2.5-2.7