We are assuming that I and J are of same dimensions. Also, we assume that in I+J corresponding pixel values of each image gets added directly. There is no after-scaling to compensate when the intensities go out of range.

Let I, J, K be the random variables for these three images respectively. Now, for each location

$$\begin{split} \operatorname{Prob}_K(k) &= \sum_{i=-\infty}^\infty \sum_{j=-\infty}^\infty \mathbbm{1}_{i+j=k} \cdot \operatorname{Prob}_{IJ}(i,j) \\ \operatorname{Prob}_K(k) &= \sum_{i=-\infty}^\infty \operatorname{Prob}_{IJ}(i,k-i) \qquad \qquad (\text{since } \mathbbm{1}_{i+j=k} = 0 \text{ for } j \neq k-i) \end{split}$$

Hence the PMF of image I+J for each location is given by $\sum_{i=-\infty}^{\infty} \mathsf{Prob}_{IJ}(i,k-i)$.

Now, we can see that there is a presence of i and k-i akin to a convolution. In fact, we can make this idea more robust if both the distributions $p_I(i)$ and $p_J(j)$ were independent. Then,

Hence the PMF of image I+J for each location is given by $p_I(i)*p_J(j)$ when the distributions are independent. This is exactly the convolution operation as studied in class.