In this part of the assignment you will experiment with computing the Bayes Score for different Bayes Net graph structures. This problem concerns two boolean variables, $X$ and $Y$.

## Question 1a

Consider the following data set of 32 observations:

| $x$ | $y$ | Number of $(X = x, Y = y)$ |
|---|---|---|
| 0 | 0 | 8 |
| 0 | 1 | 8 |
| 1 | 0 | 7 |
| 1 | 1 | 9 |

Answer the following two questions:

**Question 1a.i**. If we set parameters to their maximum-likelihood estimates, which Bayes Net achieves a higher likelihood – the one with no edges, or the one with an edge from $X$ to $Y$?

**Question 1a.ii**. Now consider computing the Bayes Score for each of two Bayes Net structures, one with no edges and the other with an edge from $X$ to $Y$. For this question, we will ignore the structure and parameter priors (i.e. we will assume they are uniform), and just focus on the integral of the likelihood over the parameter space. Compute the value of the log likelihood integrated over the parameter space:

$$\text{Score}(G) = \ln \int_{\theta_{\mathbf{G}}} P(\text{Data}|\theta_{\mathbf{G}}, G)P(\theta_{\mathbf{G}}|G)d\theta_{\mathbf{G}} \tag{1}$$

Note that this equation is equal to the Bayes Score, except we have omitted the structure prior $P(G)$ because it is the same for all graphs.

For the prior $P(\theta_{\mathbf{G}}|G)$, use Beta(1, 1) for the prior on all CPTs. Note that this prior does not change how the likelihood varies with $\theta_G$ but does introduce a normalization constant that *may* differ between the two graphs. See `https://en.wikipedia.org/wiki/Beta_distribution` to get the constant. Also, for the likelihood you do *not* need to include the number of orders in which the sufficient statistics in the table could have arisen – that is, the binomial likelihood in principle includes an "$n$ choose $k$" term representing all the orders that $k$ ones could have occurred in $n$ independent draws. You do not need to include this term in your likelihood.

Report the value of Equation 1 for the Bayes Net with no edges, and the Bayes Net with an edge from $X$ to $Y$. Which graph is preferred under the Bayes Score? How does this compare to the answer you got for question i?

Note: you can compute or approximate the integral any way you like. One way to quickly numerically approximate an integral like this is using Wolfram Alpha, with the `NIntegrate` command e.g.:

```
log[Integrate[xy, {x, 0, 1}, {y, 0, 1}]]
```

computes the natural log of the integral of the function $f(x, y) = xy$ over $0 \leq x, y \leq 1$. You may also find the `Gamma` function useful.

**Question 1a.iii**. Finally, perform the same analysis as in the previous question for the following data set.

| $x$ | $y$ | Number of $(X = x, Y = y)$ |
|---|---|---|
| 0 | 0 | 12 |
| 0 | 1 | 4 |
| 1 | 0 | 3 |
| 1 | 1 | 13 |

Again, compute the value of Equation 1 for the same two Bayes Net structures. Which graph is preferred for this data set, according to the Bayes Score? In a sentence, why does this preference make sense?