Question 1. Sketch the cumulative distribution function, and calculate the density function, for this continuous random variable:

```python
def rx():
    u = random.random()
    return u * (1-u)
```

[Hint. See Exercise 3.3, from lecture 2. Sketch a graph of \( u(1-u) \) as a function of \( u \). For what ranges of \( u \) is \( u(1-u) \leq y? \) What is the probability that the random variable \( U \sim U[0,1] \) lies in these ranges? ]

Question 2. We wish to implement a random variable whose cumulative distribution function \( F(x) = P(X \leq x) \) is given by the function below. Here, \( a \) and \( b \) are parameters in the range \([0,1]\).

\[
F(x) = \begin{cases} 
0 & \text{if } x < 0 \\
{bx/a} & \text{if } 0 \leq x \leq a \\
{b+(1-b)(x-a)/(1-a)} & \text{if } a < x \leq 1 \\
1 & \text{if } x > 1 
\end{cases}
\]

[Hint. See slide 10 from lecture 2. Also see the solution to mock exam question 1(b), in a video posted on Moodle, which suggests inventing a “mixture of uniforms”. Try answering the question first for parameters \( a = 1/2, b = 1/4 \), and after that go on to the general case.]

Question 3. Given a dataset \((x_1, \ldots, x_n)\), we wish to fit a Poisson distribution. This is a discrete random variable with a single parameter \( \lambda > 0 \), called the rate, and

\[
\Pr(x | \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x \in \{0, 1, 2, \ldots\}.
\]

Show that the maximum likelihood estimator for \( \lambda \) is \( \hat{\lambda} = n^{-1} \sum_{i=1}^{n} x_i \). [Hint. This is a question about learning generative models. See section 1.5 exercise 1.7.]

Question 4. Given a dataset \([3,2,8,1,5,0,8]\), we wish to fit a Poisson distribution. Give code to achieve this fit, using scipy.optimize.fmin. [Hint. See section 1.2 exercise 1.4. Also, if you use numpy, watch out for which variables in your code are vectors and which are scalars.]

Question 5. Given a dataset \((x_1, \ldots, x_n)\), we wish to fit the Uniform\([0, \theta]\) distribution, where \( \theta \) is unknown. By writing the density with explicit boundaries,

\[
\Pr(x | \theta) = \frac{1}{\theta} 1_{x \geq 0} 1_{x \leq \theta} \quad \text{for } x \in \mathbb{R},
\]

show that the maximum likelihood estimator is \( \hat{\theta} = \max_i x_i \).

[Hint. In any question where the range of the random variable depends on unknown parameters, it’s a good idea to include the boundaries explicitly in your density function, using an indicator function. See lecture 2 slides 10–11. A neat thing about indicator functions is that

\[
1_{\xi \geq a} \times 1_{\xi \geq b} = 1_{\xi \geq a} \text{ and } 1_{\xi \geq b} = 1_{\xi \geq \max(a,b)}.
\]

Use indicator functions, including this equation, in your answer. It will help make sure you’re not missing any corner cases.]
Question 6 (A/B testing). Your company has two systems which it wishes to compare, \(A\) and \(B\). It has asked you to compare the two, on the basis of performance measurements \((x_1, \ldots, x_m)\) from system \(A\) and \((y_1, \ldots, y_n)\) from system \(B\). Any fool using Excel can just compare the averages, 
\[
\bar{x} = \frac{1}{m-1} \sum_{i=1}^{m} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n-1} \sum_{i=1}^{n} y_i,
\]
but you are cleverer than that and you will harness the power of Machine Learning.

Suppose the \(x_i\) are drawn from \(X \sim \text{Normal}(\mu, \sigma^2)\), and the \(y_i\) are drawn from \(Y \sim \text{Normal}(\mu + \delta, \sigma^2)\), and all the samples are independent, and \(\mu, \delta, \text{and} \sigma\) are unknown. Find maximum likelihood estimators for the three unknown parameters. \([\text{Hint. See exercise 1.8. When you do maximum likelihood estimation, you are optimizing } \log \text{lik}(\text{params}|\text{data}), \text{and the data should include absolutely all data that can shed light on the params. Don’t estimate } \sigma \text{ from the } x_i \text{ alone—you should find a way to estimate it from both the } x_i \text{ and the } y_i, \text{ since both of them are informative about it.}]\]

Question 7. Let \(x_i\) be the population of city \(i\), and let \(y_i\) be the number of crimes reported. Consider the model \(Y_i \sim \text{Poisson}(\lambda x_i)\), where \(\lambda > 0\) is an unknown parameter. Find the maximum likelihood estimator \(\hat{\lambda}\). \([\text{Hint. This is a question about supervised learning. See section 1.6 exercise 1.11.}]\]

Question 8. We wish to fit a piecewise linear line to a dataset, as shown below. The inflection point is given, and we wish to estimate the slopes and intercepts. Explain how to achieve this using a linear modelling approach.

\[\text{Hint. See sections 2.2.1 and 2.2.2. Your model should represent a continuous line with an inflection point, not two separate lines. As a sanity check, you could implement your model equation and plot it. The code below illustrates a model that fails the sanity check:}
\]
\[
def \text{pred}(x, m_1,c_1,m_2,c_2, \text{inflection}_x=3):
    e = \text{numpy.where}(x <= \text{inflection}_x, 1, 0)
    \text{return } e*(m_1*x + c_1) + (1-e)*(m_2*x+c_2)
\]
\[
x = \text{numpy.linspace}(0.5,1000)
\text{plt.plot}(x, \text{pred}(x, m_1=0.5,c_1=0,m_2=1,c_2=2))
\]

Question 9. For the climate data from section 2.2.5 of lecture notes, we proposed the model
\[
\text{temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t
\]
in which the \(+\gamma t\) term asserts that temperatures are increasing at a constant rate. We might suspect though that temperatures are increasing non-linearly. To test this, we can create a non-numerical feature out of \(t\) by
\[
\text{u} = \text{'decade_' + str(math.floor(t/10)) + '0s'}
\]
(which gives us values like ’decade_1980s’, ’decade_1990s’, etc.) and fit the model
\[
\text{temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma u.
\]
Write this as a linear model, and give code to fit it. \([\text{See section 2.2.2. You should explain what the feature vectors are, then give a one-line command to estimate the parameters.}]\]

Question 10. I have two feature vectors
\[
\text{gender} = [f, f, f, f, m, m], \quad \text{eth} = [a, a, b, w, a, b, b]
\]
and I one-hot encode them as

\[ g_1 = [1, 1, 1, 0, 0, 0] \quad e_1 = [1, 1, 0, 1, 0, 0] \]
\[ g_2 = [0, 0, 0, 1, 1, 1] \quad e_2 = [0, 0, 1, 0, 1, 1] \]
\[ e_3 = [0, 0, 0, 1, 0, 0] \]

Are these five vectors \( \{g_1, g_2, e_1, e_2, e_3\} \) linearly independent? If not, find a linearly independent set of vectors that spans the same feature space. [Hint. See section 2.5 exercise 2.3.]

**Question 11.** For the police stop-and-search dataset in section 2.5 example 2.4, we wish to investigate intersectionality in police bias. We propose the linear model

\[ 1[\text{outcome}=\text{find}] \approx \alpha_{\text{gender}} + \beta_{\text{eth}}. \]

Write this as a linear model using one-hot coding. Are the parameters identifiable? If not, rewrite the model so they are, and interpret the parameters of your model. [Hint. Section 2.5 example 2.4.]
These questions are not intended for supervision (unless your supervisor directs you otherwise). Some of them are longer form exam-style questions, which you can use for revision. Others, labelled *, ask you to think outside the box.

**Question 12 (Cardinality estimation).**

(a) Let $T$ be the maximum of $m$ independent Uniform[0, 1] random variables. Show that $P(T \leq t) = t^m$. Find the density function $P_T(t)$. Hint. For two independent random variables $U$ and $V$,

$$P(\max(U, V) \leq x) = P(U \leq x \text{ and } V \leq x) = P(U \leq x) P(V \leq x).$$

(b) A common task in data processing is counting the number of unique items in a collection. When the collection is too large to hold in memory, we may wish to use fast approximation methods, such as the following: Given a collection of items $a_1, a_2, \ldots$, compute the hash of each item $x_1 = h(a_1), x_2 = h(a_2), \ldots$, then compute $t = \max_i x_i$.

If the hash function is well designed, then each $x_i$ can be treated as if it were sampled from Uniform[0, 1], and unequal items will yield independent samples.

The more unique items there are, the larger we expect $t$ to be. Given an observed value $t$, find the maximum likelihood estimator for the number of unique items. [Hint. This is about finding the mle from a single observation, as in exercise 1.1.]

http://blog.notdot.net/2012/09/Dam-Cool-Algorithms-Cardinality-Estimation

**Question 13*. Sketch the cumulative distribution functions for these two random variables. Are they discrete or continuous?

```python
def rx():
    u = random.random()
    return 1/u

def ry():
    u2 = random.random()
    return rx() + math.floor(u2)
```

[Hint. For intuition, use simulation. Generate say 10,000 samples, and plot a histogram, then a plot of “how many are $\leq x$" as a function of $x$.]

**Question 14.** A point lightsource at coordinates (0, 1) sends out a ray of light at an angle $\Theta$ chosen uniformly in $[-\pi/2, \pi/2]$. Let $X$ be the point where the ray intersects the horizontal line through the origin. What is the density of $X$? [Hint. See exercise 3.3, from lecture 2.]

Note: This random variable is known as the Cauchy distribution. It is unusual in that it has no mean.

**Question 15.** As an alternative to the model from question 9, we might suspect that temperatures are increasing linearly up to 1980, and that they are increasing linearly at a different rate from 1980 onwards. Devise a linear model to express this, using your answer to question 8, and fit it. Plot your fit. [Hint. Sample code for plotting a fit is shown in section 2.2.4.]

**Question 16 (A/B testing).** The dataset for question 6 has been presented to you as a spreadsheet with $m+n$ rows and two columns, one column measurement containing $(x_1, \ldots, x_m, y_1, \ldots, y_n)$,
and another column system whose entries are either A or B indicating which system the measurement came from.

Write the probabilistic model from question 6 as a linear model, with coefficients $\mu$ and $\delta$. Explain what your feature vectors are. [Hint. Use the approach of section 2.4.]

**Question 17.** Here are two different models for the climate data:

\[ \text{temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma t \]

and

\[ \text{temp} \approx \alpha + \beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t) + \gamma (t - 2000). \]

The first model produces a fitted value $\alpha = -63.9^\circ C$ and the second model produces a fitted value $\alpha = 10.5^\circ C$. Why the difference? Which is correct? [The answer is in section 2.2.5. But try to answer yourself, before looking it up.]

**Question 18 (Heteroscedasticity).** We are given a dataset\(^1\) with predictor $x$ and label $y$ and we fit the linear model

\[ y_i \approx \alpha + \beta_1 x_i + \gamma x_i^2. \]

After fitting the model using the least squares estimation, we plot the residuals $\varepsilon_i = y_i - (\hat{\alpha} + \hat{\beta}_1 x_i + \hat{\gamma} x_i^2)$.

(a) Describe what you would expect to see in the residual plot, if the assumptions behind linear regression are correct.

(b) This residual plot suggests that perhaps $\varepsilon_i \sim \text{Normal}(0, (\sigma x_i)^2)$ where $\sigma$ is an unknown parameter. Assuming this is the case, give pseudocode to find the maximum likelihood estimators for $\alpha$, $\beta_1$, and $\gamma$.

[Hint. This question is asking you to reason about a custom probability model, in the style of section 2.4. A model with unequal variances is called ‘heteroscedastic’.]

**Question 19.** Let $(F_1, F_2, F_3, \ldots) = (1, 1, 2, 3, \ldots)$ be the Fibonacci numbers, $F_n = F_{n-1} + F_{n-2}$. Define the vectors $f$, $f_1$, $f_2$, and $f_3$ by

\[
\begin{align*}
    f &= [F_4, F_5, F_6, \ldots, F_{m+3}] \\
    f_1 &= [F_3, F_4, F_5, \ldots, F_{m+2}] \\
    f_2 &= [F_2, F_3, F_4, \ldots, F_{m+1}] \\
    f_3 &= [F_1, F_2, F_3, \ldots, F_m]
\end{align*}
\]

for some large value of $m$. If you were to fit the linear model

\[ f \approx \alpha + \beta_1 f_1 + \beta_2 f_2 \]

what parameters would you expect? What about the linear model

\[ f \approx \alpha + \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3? \]

[Hint. Are the feature vectors linearly independent?]

\(^1\)https://teachingfiles.blob.core.windows.net/datasets/heteroscedasticity.csv