(1: Getting Started in MATLAB): Suppose that \( x \) is a random variable that is uniformly distributed over the interval \([1, 2]\). Thus, the density function, \( p(x) \), is
\[
p(x) = (1)I(1 < x < 2),
\]
and \( I(\cdot) \) denotes a standard indicator function, which equals one of the event is true and zero otherwise.

Let \( y = x^2 \). [Note that \( y \in (1, 4) \), given that \( x \in (1, 2) \).]

Derive the density function for \( y \), \( p(y) \). Hint: There are a variety of ways to do this, but perhaps the easiest is to use a simple change of variables (remembering, of course, to include the Jacobian term). An alternative is the c.d.f. method. That is, first derive \( F_y(c) = \Pr(y \leq c) \) for \( c \in (1, 4) \). Once you have this, differentiate the c.d.f. with respect to \( c \) to obtain the density function. Verify that this is a proper density function (i.e., it integrates to one over its support), and also calculate the mean of \( y \).

Plot \( p(y) \) using MATLAB. To do this, first create a grid of points over \((1, 4)\), [Note: the “linspace” command is helpful for this] and then calculate the density at each of these points. The “plot” command in MATLAB can be used for plotting.

Use the “trapz” command (for information on this, type “help trapz” in MATLAB) to numerically calculate the area under the density (it should be close to one!) and the mean (it should be close to the mean you derived analytically above).

(2) Consider the random effects approach to the panel data model:
\[
y_{it} = x_{it}\beta + a_i + u_{it},
\]
where \( a_i \) is a random effect generating correlation in outcomes over time for the \( i^{th} \) unit. We denote \( \text{Var}(a_i) = \sigma_a^2 \), \( \text{Var}(u_{it}) = \sigma_u^2 \). We also continue to make all of the assumptions described in your class notes regarding correlation between the random effects \( a \) and the regression error \( u \) (i.e., the \( u \)'s are independent of one another and are independent of \( a \)). Let \( v_{it} \) denote the composite error: \( v_{it} = a_i + u_{it} \).
In class, we showed that the GLS estimator of $\beta$ in (1) is obtained by running OLS using the quasi-demeaned data:

$$y_{it} - \lambda \overline{y}_i = (x_{it} - \lambda \overline{x}_i)\beta + [v_{it} - \lambda \overline{v}_i],$$  \hspace{1cm} (2)

where

$$\lambda = 1 - \left[ \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2} \right]^{1/2}.$$  

At first glance, it is not obvious that this transformation does what it is reported to do: to transform the covariance matrix of the original representation of the model in (1) to a covariance matrix that falls under the assumptions of the classical linear regression model. That is, once transformed in this fashion, the regression errors in (2) are homoscedastic and serially uncorrelated.

Let $\epsilon_{it} = v_{it} - \lambda \overline{v}_i$. Show that $\epsilon_{it}$ are mean zero, homoscedastic and serially uncorrelated.

[Hint: To establish homoscedasticity, derive $\text{Var}(\epsilon_{it})$. To establish serial uncorrelation, evaluate $\text{Cov}(\epsilon_{ik}, \epsilon_{ij})$. Finally, at some point in the derivations, I found it useful to write $\lambda^2 - 2\lambda = \lambda^2 - 2\lambda + 1 - 1 = (\lambda - 1)^2 - 1$.]

(3) Consider the random effects balanced panel data model, which has been averaged over time:

$$\overline{y}_i = \beta_0 + \beta_1 \overline{x}_i + a_i + \overline{u}_i, \quad i = 1, 2, \ldots, n.$$  

Continue to assume that $\overline{u}_i$ is uncorrelated with $\overline{x}_i$, a scalar, but $\text{Cov}(x_{it}, a_i) = \sigma_{xa}$ for all $i$ and $t$.

(3a) Show that the probability limit (or plim) of the OLS estimator of $\beta_1$ is

$$\text{plim}(\hat{\beta}_1) = \beta_1 + \frac{\sigma_{xa}}{\text{Var}(\overline{x}_i)}.$$  

(3b) Now, suppose that $x_{it}$ is uncorrelated over time with $\text{Var}(x_{it}) = \sigma_x^2$. Using your result in (3a), how is the inconsistency of the OLS estimator affected by the number of time periods $T$?