1. Suppose that X and Y are continuous random variables with CDF and conditional pdf, defined as:

$$F_X(x) = \begin{cases} 0 & x \le 0 \\ x^2 & 0 < x \le 1 \\ 1 & 1 < x \end{cases}$$

$$f_{Y|X}(y|x) = \begin{cases} 0 & y \le 0 \\ 3y^2/x^3 & 0 < y \le x \\ 0 & x < y \end{cases}$$

   (a) Find the joint probability density function, $f_{X,Y}(x, y)$.
   (b) Find the marginal probability density function, $f_Y(y)$.
   (c) Find the conditional probability density function, $f_{X|Y}(x|y)$.
   (d) Find the conditional expectation of X, given the value y.

2. Let $X_i, i = 1, 2, 3, ...$ be i.i.d. with exponential distribution.

   (a) Let $S_n = \sum_{i=1}^n X_i / \sqrt{n \sum_{i=1}^n X_i^2}$. Show that S_n converges in probability, and derive its limit.
   (b) $T_n = \sqrt{n} \sum_{i=1}^n (X_i - \theta) / \sum_{i=1}^n X_i^2$. Show that T_n converges in distribution, and derive its limiting distribution.
   (c) Let $U_n = \sum_{i=1}^n X_i$. Find the moment generating function of X_i.
   (d) Let $U_n = \sum_{i=1}^n X_i$. Find the moment generating function of U_n (for fixed value of n).
   (e) Using the previous part, find $E(U_n^3)$.
   (f) Let $V = \frac{X_2}{X_1}$. Derive the CDF $F_V(v)$.

3. Let y_{ti} be independent exponential random variables where the mean of y_{ti} is θ_t, $t = 1, \ldots, T$, and $i = 1, \ldots, n$.

   (a) Derive the MLE of $\theta_1, \ldots, \theta_T$.
   (b) Derive the asymptotic size α LRT of the hypothesis that all the θ_t are equal.
   (c) Using the pivotal quantity approach find a $1 - \alpha$ two-sided confidence interval for $\tau = \theta_1/\theta_2$.
   (d) Use the confidence interval you found in 3c to find an exact test of the hypothesis $\theta_1 = \theta_2$.
   (e) Let the prior distribution of the $\theta_1, \ldots, \theta_T$ be independent inverted-gamma(1,1) distributions. The pdf of an inverted-gamma distribution is

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{-(\alpha+1)} e^{-1/(\beta x)}; \quad x > 0.$$

   Find a $1 - \alpha$ credible interval for θ_1.

4. In the innovation campus greenhouse at University of Nebraska - Lincoln a research group planted and evaluated 100 sorghum genotypes (lines). They wanted to see how different factors influence the height of the plants during the growing season. They had 10 large bins that was able to hold 100 pots each. The bins were placed in different locations of the greenhouse. In each bin they had all of the lines planted. In each pot they planted 6 seeds (which resulted in 6 plants per pot).

   Throughout the growing season - after emergence of the plants - they collected image data for 20 consecutive days, and a total of 2,000 images were collected using 10 cameras (10 fields$\times$ 20 days $\times$ 10 cameras).

   The image processing step was a collaborative effort between several universities where undergraduate students were randomly assigned to process some images. The images were processed in batches. They had 200 batches (field by day combinations) of 10 images (taken with the 10 cameras). Each batch of images was processed by two students, so a total of 400 students were involved in this step of the research. The students had to determine the height of each plant by using an image processing software.

   Our response is the height of the plants at day 20 determined by the students using the image processing software. We are considering the following model.

   $$y_{blps} = \mu + \beta_b + \tau_l + a_{bl} + b_{blp} + c_{bs} + e_{blps}$$

   where $a_{bl} \sim N(0, \sigma_a^2), b_{blp} \sim N(0, \sigma_b^2), c_{bs} \sim N(0, \sigma_c^2), e_{blps} \sim N(0, \sigma_e^2)$.

   (b stands for bin, l stands for line, p stands for plant and s stands for student.)
   We also assume that all of the random effects and errors are independent.

   (a) Let $\bar{y}_{.l..}$ be the average of all the approximated plant heights on day 20. Find the expected value of $\bar{y}_{.1..}$ in terms of the model parameters.

   (b) Find the variance of $\bar{y}_{.1..}$ in terms of the model parameters.

   (c) Find the variance of $\bar{y}_{.1..} - \bar{y}_{.2..}$ in terms of the model parameters.

   (d) The model described above was fit to the approximated plant heights for day 20. A portion of R code and output is provided in the Appendix. Determine the value of a t-statistic for testing $H_0 : \tau_9 = \tau_{10}$.

   (e) Determine the degrees of freedom for the test statistic in part (d).

   (f) The REML method was used to estimate the variance components. The REML method involves a likelihood function for observations known as error contrasts. How many error contrasts are involved in the REML likelihood function for the fit of the model specified above to the approximated heights collected for day 20?

   (g) Suppose some students evaluate the images differently than others. Explain how this type of variability is accounted for in the model or how should we modify the model to account for such variability.

5. Lets consider a linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, where $\mathbf{y}$ has dimension of $n \times 1$, $\mathbf{X}$ has dimension of $n \times p$ with rank of r. β has dimension of $\mathbf{p} \times \mathbf{1}$, and is a vector of unknown parameters, and $e \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Also, suppose $\mathbf{c}$ is a known, p-dimensional vector.

   (a) What must be true of $\mathbf{c}$ in order for $\mathbf{c}'\beta$ to be estimable?

   (b) Suppose $\mathbf{c}'\beta$ is estimable. Provide an expression for the BLUE OF $\mathbf{c}'\beta$ in terms of $\mathbf{c}, \mathbf{X}, \mathbf{y}$.

   (c) Suppose $\mathbf{c}'\beta$ is estimable. Find the distribution of the BLUE of $\mathbf{c}'\beta$.

   (d) Suppose $\mathbf{c}'\beta$ is estimable. Provide and expression for the 95% confidence interval for $\mathbf{c}'\beta$.

   (e) Show that the confidence interval provided in part (d) has coverage probability 0.95.

6. State the Gauss Markov theorem. Be specific.

Appendix

Partial R code and Output

>M=lmer(y~field+lines+(1|a)+(1|b)+(1|c))

>summary(M)

Linear mixed model fit by REML [‘lmerMod’]

Formula: y~field+lines+(1|a)+(1|b)+(1|c)

REML criterion at convergence: 51937.8

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	152.35029	1.09520	139.107
Field2	-8.91885	1.41863	-6.287
Field3	-6.63456	1.41863	-4.677
Field4	0.06360	1.41863	0.045
Field5	5.22145	1.41863	3.681
Field6	-1.62157	1.41863	-1.143
Field7	-6.10325	1.41863	-4.302
Field8	13.51825	1.41863	9.529
Field9	2.24084	1.41863	1.580
Field10	2.32538	1.41863	1.639
Line2	13.01750	0.62477	20.836
Line3	-4.02125	0.62477	-6.436
Line4	6.06067	0.62477	9.701
Line5	23.39275	0.62477	37.442
Line6	0.38508	0.62477	0.616
Line7	-10.55892	0.62477	-16.901
Line8	-6.23683	0.62477	-9.983
Line9	11.99942	0.62477	19.206
Line10	-4.50458	0.62477	-7.210

[ADDITIONAL LINES OMITTED]