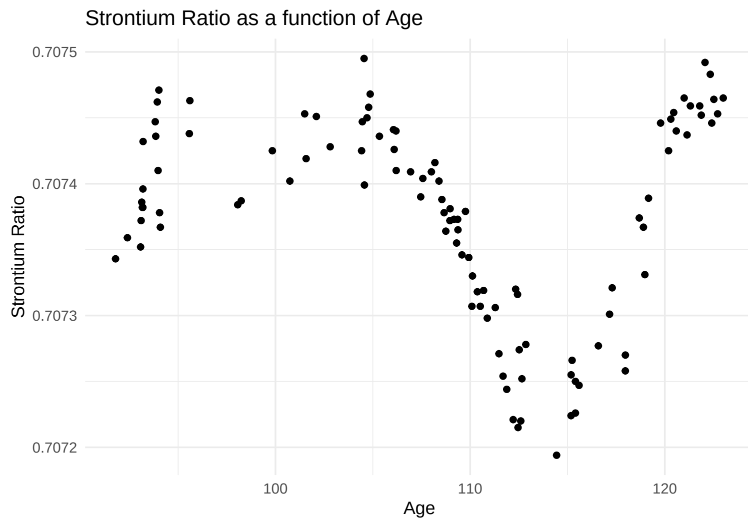
Problem 1. In the following problem we will consider smoothing and inference for the fossil data available at https://donatello-telesca.squarespace.com/s/fossil.txt. These data consist of 106 measurements of ratios of strontium isotopes found in fossil shells and their age. We are interested in describing isotope ratios as a function of the fossil's age.

Let y =ratio and x =age of strontium isotopes. Let $x_0 = \min(x)$ and $x_1 = \max(x)$. Given a set of equally spaced interior knots $x_0 + \delta = \eta_1 < \ldots < \eta_{20} = x_1 - \delta$, $\delta > 0$, consider the following model

$$y_i = f(x_i) + \epsilon_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \sum_{j=1}^{20} b_j (x_i - \eta_j)_+^3 + \epsilon_i, \tag{1}$$

with $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$, and $Cov(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$.

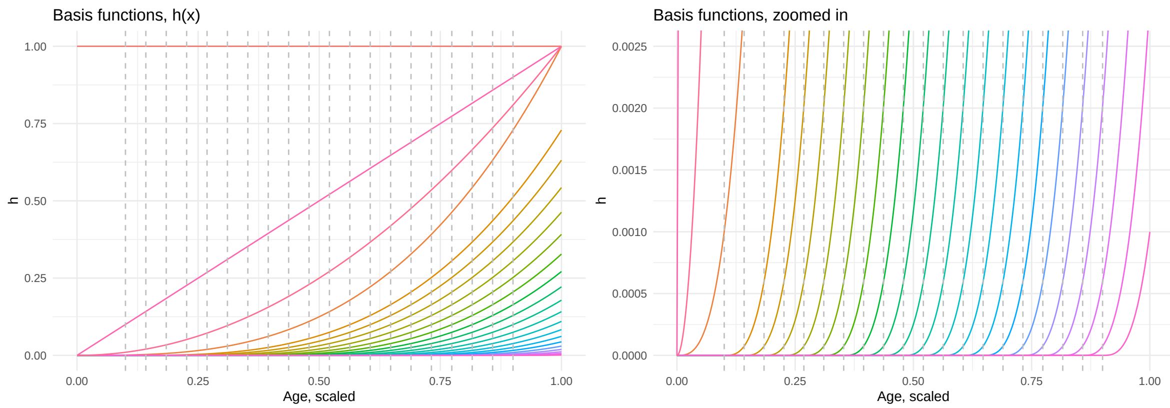


a. Plot the basis functions defining spanning the space of $f(\cdot)$, within the domain of x. You may find it useful to rescale the domain of x to be $D = [0, 1]$.

Solution. Here I've plotted the basis functions on the left with the knots denoted by the dashed vertical grey lines. On the right, I've plotted the same functions zoomed in on the y-axis to show how each knot has a cubic function starting at the knot value.

```
# add a new variable `age_scaled` that is age scaled between 0 and 1
# arrange by `age_scaled`
fossil <- mutate(
  fossil,
  age_scaled = rescale(age)
)

# create a vector of 106 timepoints between 0 and 1 to use for the basis
# functions
timepoints <- seq(0, 1, length.out = 1e3)

# create the basis functions
# there should terms for: a constant, linear, quadratic and cubic, as well
# as cubic functions starting at each knot
C <- data.frame(
  constant = 1,
  linear = timepoints,
  quadratic = timepoints^2,
  cubic = timepoints^3
```

```
) %>%
  cbind(
    map_dfc(
      seq(0.1, 0.9, length.out = 20),
      ~ ifelse(timepoints < .x, 0, (timepoints - .x)^3)
    ) %>%
      setNames(
        map_chr(
          1:20,
          ~ paste("knot", str_pad(.x, width = 2, pad = "0"), sep = "_")
        )
      )
  )
```



b. Let $y \in \mathbb{R}^{106}$, and write the model in (1) in vector form, as $y = X\beta + Zb + \epsilon = C\gamma + \epsilon$. Describe the form of the matrices X, Z, and C.

Solution. If we stack the strontium ratio observations into the vector $y \in \mathbb{R}^{106}$, then, using the scaled age values, the matrix X is the 106×4 matrix with the columns for the intercept, age, age squared, and age cubed

$$X = \begin{bmatrix} 1 & \text{age}_1 & \text{age}_1^2 & \text{age}_1^3 \\ 1 & \text{age}_2 & \text{age}_2^2 & \text{age}_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{age}_{105} & \text{age}_{105}^2 & \text{age}_{105}^3 \\ 1 & \text{age}_{106} & \text{age}_{106}^2 & \text{age}_{106}^3 \end{bmatrix}$$

The matrix Z is the 106×20 matrix with columns for age minus η_1 cubed, age - η_2 cubed, $\ldots$, age - η_{20} cubed

$$Z = \begin{bmatrix} (\text{age}_1 - \eta_1)^3 I(\text{age}_1 > \eta_1) & (\text{age}_1 - \eta_2)^3 I(\text{age}_1 > \eta_2) & \cdots & (\text{age}_1 - \eta_{20})^3 I(\text{age}_1 > \eta_{20}) \\ (\text{age}_2 - \eta_1)^3 I(\text{age}_2 > \eta_1) & (\text{age}_2 - \eta_2)^3 I(\text{age}_2 > \eta_2) & \cdots & (\text{age}_2 - \eta_{20})^3 I(\text{age}_2 > \eta_{20}) \\ \vdots & \vdots & & \vdots \\ (\text{age}_{105} - \eta_1)^3 I(\text{age}_{105} > \eta_1) & (\text{age}_{105} - \eta_2)^3 I(\text{age}_{105} > \eta_2) & \cdots & (\text{age}_{105} - \eta_{20})^3 I(\text{age}_{105} > \eta_{20}) \\ (\text{age}_{106} - \eta_1)^3 I(\text{age}_{106} > \eta_1) & (\text{age}_{106} - \eta_2)^3 I(\text{age}_{106} > \eta_2) & \cdots & (\text{age}_{106} - \eta_{20})^3 I(\text{age}_{106} > \eta_{20}) \end{bmatrix}$$

The (106×24) matrix C is formed by concatenating the columns of X and Z

$$C = \begin{bmatrix} 1 & \text{age}_1 & \text{age}_1^2 & \text{age}_1^3 & (\text{age}_1 - \eta_1)^3 I(\text{age}_1 > \eta_1) & \cdots & (\text{age}_1 - \eta_{20})^3 I(\text{age}_1 > \eta_{20}) \\ 1 & \text{age}_2 & \text{age}_2^2 & \text{age}_2^3 & (\text{age}_2 - \eta_1)^3 I(\text{age}_2 > \eta_1) & \cdots & (\text{age}_2 - \eta_{20})^3 I(\text{age}_2 > \eta_{20}) \\ \vdots & \vdots & \vdots & \vdots & & & \\ 1 & \text{age}_{105} & \text{age}_{105}^2 & \text{age}_{105}^3 & (\text{age}_{105} - \eta_1)^3 I(\text{age}_{105} > \eta_1) & \cdots & (\text{age}_{105} - \eta_{20})^3 I(\text{age}_{105} > \eta_{20}) \\ 1 & \text{age}_{106} & \text{age}_{106}^2 & \text{age}_{106}^3 & (\text{age}_{106} - \eta_1)^3 I(\text{age}_{106} > \eta_1) & \cdots & (\text{age}_{106} - \eta_{20})^3 I(\text{age}_{106} > \eta_{20}) \end{bmatrix}$$

c. Let $D = \text{diag}(0, 0, 0, 0, 1, \ldots, 1)_{24\times 24}$ be a penalization matrix. For any $\lambda > 0$, describe how $\hat{f}_\lambda(x)$ is estimated, as the solution to the PLS problem.

$$||y - C\gamma||^2 + \lambda\gamma' D\gamma.$$

Solution. Given a set of J spline basis functions $\{C_1, \ldots, C_J\}$, we want to approximate $\hat{f}_\lambda(x)$ with $\hat{\gamma}_\lambda C(x)$. The penalized regression spline $\hat{f}_\lambda$ is obtained by minimizing

$$||y - C\gamma||^2 + \lambda\gamma' D\gamma$$

which gives the solution

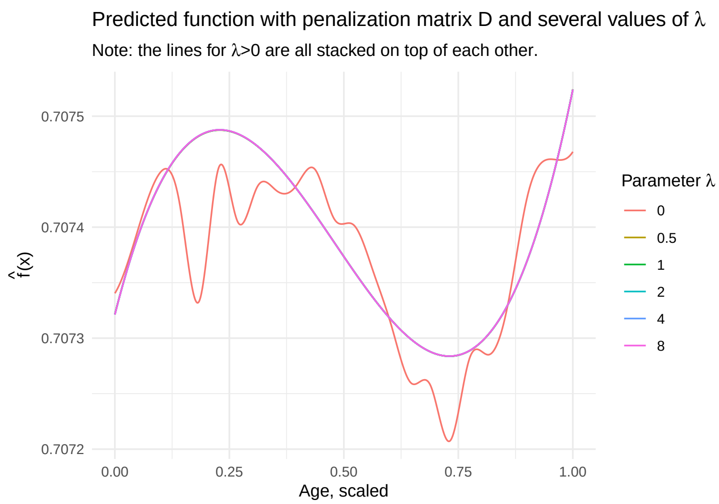
$$\hat{\gamma}_\lambda = (C^{\mathrm{T}} C + \lambda D)^{-1} C^{\mathrm{T}} Y.$$

d. Given the structure of D, describe how the formula above acts on $\hat{f}(x)$, to regularize its flexibility. Corroborate your analysis by plotting $\hat{f}_\lambda(x)$, for $\lambda = 0, 0.5, 1, 2, 4, 8$.

Solution. Given the structure of D, the formula above will penalize large values of the coefficients of the piecewise cubic functions more as λ increases, shrinking their coefficients to zero.

```r
# create D matrix as described above
D <- diag(x = c(rep(0, 4), rep(1, 20)))

# create a named vector of lambdas to use in purrr::map
lambda <- c(
  "lambda_0"  = 0,
  "lambda_05" = 0.5,
  "lambda_1"  = 1,
  "lambda_2"  = 2,
  "lambda_4"  = 4,
  "lambda_8"  = 8
)

C_tilde <- data.frame(
  constant = 1,
  linear = fossil$age_scaled,
  quadratic = fossil$age_scaled^2,
  cubic = fossil$age_scaled^3
) %>%
  cbind(
    map_dfc(
      seq(0.1, 0.9, length.out = 20),
      ~ ifelse(fossil$age_scaled < .x, 0, (fossil$age_scaled - .x)^3)
    ) %>%
      setNames(
        map_chr(
          1:20,
          ~ paste("knot", str_pad(.x, width = 2, pad = "0"), sep = "_")
        )
      )
  )
```

```r
# calculate gamma_hat for all values in lambda vector
gamma_hat <- map(
  lambda,
  ~ solve(
    t(as.matrix(C_tilde)) %*% as.matrix(C_tilde) + .x * D,
    t(as.matrix(C_tilde)) %*% fossil$strontium.ratio,
    tol = 4e-17
  )
)

# calculate predicted function using gamma_hat
f_hat_plot <- map_dfc(
  gamma_hat,
  ~ as.matrix(C) %*% .x
)
```

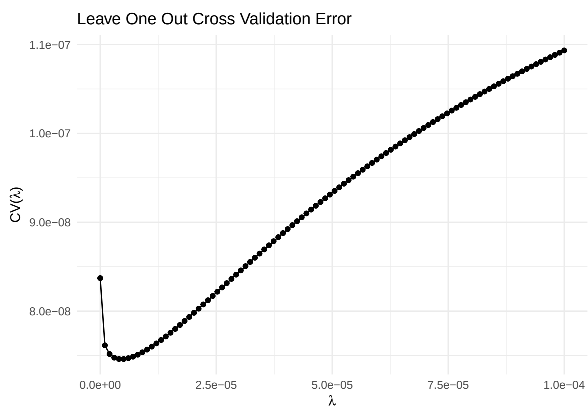


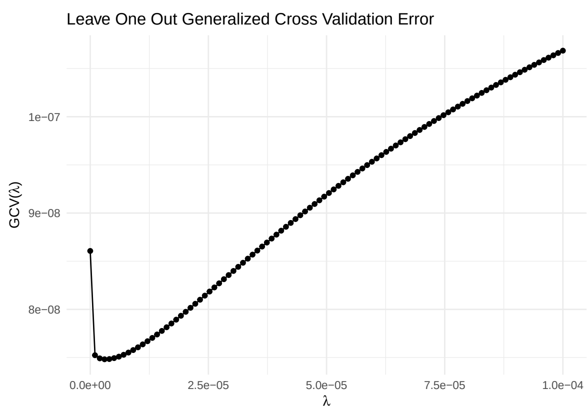
e. Find the optimal values of λ minimizing the CV and GCV errors.

Solution.

```r
# create a grid of lambda values to find the optimal value
lambda <- seq(0, 0.0001, length.out = 100)

# calculate the smoothing matrix S
smooth_mat_tilde <- map(
  lambda,
  ~ as.matrix(C_tilde) %*%
    solve(
      t(as.matrix(C_tilde)) %*% as.matrix(C_tilde) + .x * D,
      t(as.matrix(C_tilde)),
      tol = 4e-17
    )
)

# calculate the predicted function values
f_hat <- map_dfc(smooth_mat_tilde, ~ .x %*% fossil$strontium.ratio)

# calculate the leave one out cross validation error for each value of
```

```r
# lambda
cvs <- map_dfr(
  1:100,
  function(x) {
    cv <- 0
    for (i in 1:106) {
      cv <- cv + ((fossil$strontium.ratio[i] - f_hat[i, x]) /
                    (1 - smooth_mat_tilde[[x]][i, i]))^2
    }
    data.frame("lambda_val" = lambda[x], "cv" = as.numeric(cv))
  }
)

# calculate the leave one out generalized cross validation error for each
# value of lambda
gcvs <- map_dfr(
  1:100,
  function(x) {
    gcv <- 0
    for (i in 1:106) {
      gcv <- gcv + ((fossil$strontium.ratio[i] - f_hat[i, x]) /
                      (1 - (1 / 106) * sum(diag(smooth_mat_tilde[[x]]))))^2
    }
    data.frame("lambda_val" = lambda[x], "gcv" = as.numeric(gcv))
  }
)
```





The optimal value of $\lambda_{\text{CV}} = 5.0505051 \times 10^{-6}$ and $\lambda_{\text{GCV}} = 3.030303 \times 10^{-6}$.

f. Show that the $\hat{\gamma}$ obtained via PLS is equivalent to ML and BLUP estimates obtained through a mixed effects model representation of (1).

Solution. The linear mixed effects model representation of (1) is

$$\begin{gathered} Y = X\beta + Zb, \\ b \sim \text{N}(0, \sigma_b^2 I_J), \\ e \sim \text{N}(0, \sigma_\epsilon^2 I_N), \end{gathered}$$

and $cov(b, e) = 0$. Conditional on b, $Y|b \sim \text{N}(X\beta + Zb, \sigma_\epsilon^2 I_N)$. However, marginalizing Y by integrating out b gives $Y \sim \text{N}(X\beta, \sigma_\epsilon^2 I_N + \sigma_b^2 ZZ^T)$, which creates a dependence structure for Y. We can estimate β using maximum likelihood which gives the generalized least squares estimate

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

where $V = \sigma_\epsilon^2 I_N + \sigma_b^2 Z Z^T$. To estimate b, we use the best linear unbiased predictor (BLUP), which takes the expectation of b given the data. Thus we need to find the joint distribution of b and Y

$$\begin{aligned} f(b,Y) &= f(Y|b)f(b) \\ &= \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\{-\frac{1}{2}(Y - X\beta - Zb)^T V^{-1}(Y - X\beta - Zb)\} \cdot \frac{1}{(2\pi)^{J/2}|\sigma_b^2|_J|^{1/2}} \exp\{-\frac{1}{2} b^T(\sigma_b^2 I_J)^{-1} b\} \end{aligned}$$

$$f(b|Y) \propto \exp\{-\frac{1}{2}(b - \sigma_b^2 Z^T V^{-1}(Y - X\beta)^T \Sigma^{-1}(b - \sigma_b^2 Z^T V^{-1}(Y - X\beta\}$$

$$\hat{b} = E[b|Y] = \sigma_b^2 Z^T V^{-1}(Y - X\hat{\beta}).$$

g. Using a mixed effects model representation estimate the smoothing parameter λ using ML and REML. Describe your work in detail.

Solution. Using the mixed effects model representation

$$Y \sim \mathrm{N}(X\beta, V(\alpha)),\ V(\alpha) = \sigma_\epsilon^2 I_n + \sigma_b^2 Z^T Z$$

the estimate of $\hat{\beta} = (X^T V(\alpha)^{-1} X)^{-1} X^T V(\alpha)^{-1} Y$ and

$$\begin{aligned} \widehat{\alpha}_{\mathrm{ML}} &= \overset{argmax}{\alpha}\ \ell_p(\alpha) = -\frac{1}{2} log|V(\alpha)| - \frac{1}{2}(Y - X\hat{\beta}(\alpha))^T V(\alpha)^{-1}(Y - X\hat{\beta}(\alpha)), \\ \widehat{\alpha}_{\mathrm{REML}} &= \overset{argmax}{\alpha}\ \ell_R(\alpha) = \ell_p(\alpha) - \frac{1}{2} log|X^T V(\alpha)^{-1} X|. \end{aligned}$$

```r
z <- as.matrix(C_tilde[, 5:24])
alpha <- expand.grid(
  error = seq(0.1, 1, length.out = 100),
  random_effect = seq(1000, 10000, length.out = 100)
) %>%
  filter(error != 0, random_effect != 0) %>%
  mutate(ratio = error/random_effect)

v_alpha <- pmap(
  alpha,
  function(error, random_effect, ratio) {
    error * diag(1, 106) + random_effect * z %*% t(z)
  }
)

x <- as.matrix(C_tilde[, 1:4])
beta_alpha <- map(
  v_alpha,
  ~ solve(t(x) %*% solve(.x) %*% x) %*% t(x) %*%
      solve(as.matrix(.x)) %*% fossil$strontium.ratio
)

ell_p <- map2_dbl(
  .x = v_alpha,
  .y = beta_alpha,
  ~ - 1/2 * log(det(.x)) -
    1/2 * t(fossil$strontium.ratio - x %*% .y) %*% solve(.x) %*%
    (fossil$strontium.ratio - x %*% .y)
```

```
)
ell_R <- ell_p - map_dbl(
  .x = v_alpha,
  ~ log(det(t(x) %*% solve(.x) %*% x))
)
```

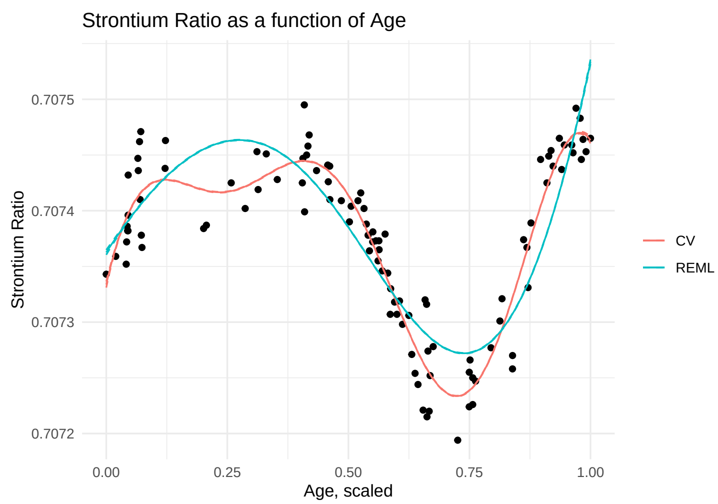
The estimated λ_{ML} is 10^{-4} and the estimated λ_{REML} is 0.001.

h. Plot the data, together with $\hat{f}(x)$, and pointwise credible bands obtained via PLS at $\hat{\lambda}_{\text{CV}}$. Compare these intervals with the ones obtained via a mixed effects model representation at $\hat{\lambda}_{\text{REML}}$.

Solution.

```
# CV
# calculate gamma_hat for lambda_cv
gamma_hat_cv <- solve(
  t(as.matrix(C_tilde)) %*% as.matrix(C_tilde) +
    cvs$lambda_val[which.min(cvs$cv)] * D,
  t(as.matrix(C_tilde)) %*% fossil$strontium.ratio,
  tol = 4e-17
)

# calculate predicted function using gamma_hat
f_hat_plot_cv <- as.matrix(C) %*% gamma_hat_cv

# estimate \hat{sigma}_\text{error} as the sd of the residuals
f_hat_cv <- as.matrix(C_tilde) %*% gamma_hat_cv
residuals_cv <- fossil$strontium.ratio - f_hat_cv
sigma_hat_cv <- sd(residuals_cv)

# estimate the upper and lower bands
smooth_mat_cv <- as.matrix(C) %*%
    solve(
      t(as.matrix(C)) %*% as.matrix(C) +
        cvs$lambda_val[which.min(cvs$cv)] * D,
      t(as.matrix(C)),
      tol = 4e-17
    )
band_upper_cv <- f_hat_plot_cv +
  2 * sigma_hat_cv * diag(t(smooth_mat_cv) %*% smooth_mat_cv)
band_lower_cv <- f_hat_plot_cv -
  2 * sigma_hat_cv * diag(t(smooth_mat_cv) %*% smooth_mat_cv)

# REML
# calculate gamma_hat for lambda_reml
gamma_hat_reml <- solve(
  t(as.matrix(C_tilde)) %*% as.matrix(C_tilde) +
    alpha[which.min(ell_R), "error"] /
    alpha[which.min(ell_R), "random_effect"] * D,
  t(as.matrix(C_tilde)) %*% fossil$strontium.ratio,
  tol = 4e-17
)

# calculate predicted function using gamma_hat
f_hat_plot_reml <- as.matrix(C) %*% gamma_hat_reml
```

```
# estimate \hat{sigma}_\text{error} as the sd of the residuals
f_hat_reml <- as.matrix(C_tilde) %*% gamma_hat_reml
residuals_reml <- fossil$strontium.ratio - f_hat_reml
sigma_hat_reml <- sd(residuals_reml)

# estimate the upper and lower bands
smooth_mat_reml <- as.matrix(C) %*%
    solve(
      t(as.matrix(C)) %*% as.matrix(C) +
        alpha[which.min(ell_R), "error"] /
        alpha[which.min(ell_R), "random_effect"] * D,
      t(as.matrix(C)),
      tol = 4e-17
    )
band_upper_reml <- f_hat_plot_reml +
  2 * sigma_hat_reml * diag(t(smooth_mat_reml) %*% smooth_mat_reml)
band_lower_reml <- f_hat_plot_reml -
  2 * sigma_hat_reml * diag(t(smooth_mat_reml) %*% smooth_mat_reml)
```



The estimated function using restricted maximum likelihood is smoother than the function estimated using cross validation.

i. Consider the problem of simultaneous credible band $[L^\star(x), U^\star(x)]$, for $f(x)$, such that

$$Pr(L^\star(x) \leq f(x) \leq U^\star(x), \forall x \in \mathcal{X}) \geq (1-\alpha).$$

Discuss possible approaches to solving this multidimensional inference problem and attempt a solution either theoretically or via Monte Carlo methods.

Solution.