Statistics I Exercises

1. The general $n \times n$ Hilbert matrix has element (i, j) given by $1/(i + j - 1)$.
   (a) Write a function which gives the $n \times n$ Hilbert matrix as its output, for any positive integer n.
   (b) Are all of the Hilbert matrices invertible?
   (c) Use `solve()` and `qr.solve()` to compute the inverse of Hilbert matrices for $n = 1$ to 10. Is there a problem?

2. Let $(x_1, x_2, x_3, x_4, x_5, x_6) = (10, 11, 12, 13, 14, 15)$. Find the coefficients of the quintic polynomial $p(x)$ for which $(p(x_1), p(x_2), p(x_3), p(x_4), p(x_5), p(x_6)) = (25, 16, 26, 19, 21, 20)$. Hint: the quintic polynomial $p(x) = a_1 + a_2x + a_3x^2 + a_4x^3 + a_5x^4 + a_6x^5$ can be viewed as the matrix product of the row vector $[1, x, x^2, x^3, x^4, x^5]$ with the column vector $[a_1, a_2, a_3, a_4, a_5, a_6]'$. Work out the matrix version of this to give $[p(x_1), p(x_2), p(x_3), p(x_4), p(x_5), p(x_6)]'$.

3. Create a 5×3 matrix X with elements drawn randomly from the uniform distribution on $[0, 1]$.
   (a) Calculate $H = X(X'X)^{-1}X'$ and the eigenvalues and eigenvectors if H.
   (b) Calculate the trace of H, and compare with the sum of the eigenvalues.
   (c) Calculate the determinant of H, and compare with the product of the eigenvalues.
   (d) Verify that the columns of X are eigenvectors of H. What are the corresponding eigenvalues?

4. Obtain the 6×6 Hilbert matrix, and compute its eigenvalues and eigenvectors. Compute the inverse of the matrix. Is there a relation between the eigenvalues of the inverse and the eigenvalues of the original matrix? Is there supposed to be a relationship?

5. Consider the following *circulant* matrix:

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.4 & 0.1 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.4 & 0.1 \end{bmatrix}$$

   (a) P is an example of a stochastic matrix. Verify that the row sums add to one.
   (b) Compute P^n for $n = 2, 3, 5, 10$. Is a pattern emerging?
   (c) Find a non-negative vector x whose elements sum to one and which satisfies $P'x = x$. Do you see any connection between P^{10} and x?

6. Consider the following matrix:

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.4 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.0 & 0.0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

   (a) P is an example of a stochastic matrix. Verify that the row sums add to one.
   (b) Compute P^n for $n = 2, 3, 5, 10$. Is a pattern emerging?

(c) Find a non-negative vector x whose elements sum to one and which satisfies $P'x = x$. Do you see any connection between P^{10} and x?

7. Let P be a permutation matrix. Show that $P^{-1} = P'$ and that P can be expressed as a product of pairwise interchanges.

8. How would you solve a partitioned linear system of the form

$$\begin{bmatrix} L_1 & O \\ B & L_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}$$

where L_1 and L_2 are non-singular lower triangular matrices, and the solution and right-hand side are partitioned accordingly?

9. Given an n-vector a, we can annihilate *all* of its entries below the kth position, provided that $\alpha_k \neq 0$, by the following transformation:

$$M_k a = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & -\mu_{k+1} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\mu_n & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ \alpha_{k+1} \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $\mu_i = \alpha_i/\alpha_k$, $i = k+1, \ldots, n$. The divisor α_k is called the pivot; a matrix of the form of M_k is called an *elementary transformation matrix.*

Establish the following properties of such matrices.

(a) M_k is a lower triangular matrix with unit main diagonal, and hence non-singular.

(b) $M_k = I - m_k e_k'$, where $m_k = [0, \ldots, 0, \mu_{k+1}, \ldots, \mu_n]'$ and e_k is the k-th Cartesian unit vector.

(c) $M_k^{-1} = I + m_k e_k'$, which mean that $M_k^{-1} = L_k$ is the same as M_k except that the signs of the multipliers are reversed.

(d) If M_l, $l > k$ is another elementary elimination matrix with m_l its vector of multipliers, then

$$M_k M_l = I - m_k e_k' - m_l e_l'.$$

10. Prove that the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

has no LU factorization, i.e., that there are no lower triangular matrix L and upper triangular matrix U such that $A = LU$.

11. Show that an $n \times n$ matrix A has rank one if and only if there are non-zero n-vectors u and v such that $A = uv'$.

12. An $n \times n$ matrix A is said to be *elementary* if it differs from the identity matrix by a matrix of rank one, i.e., if $A = I - uv'$ for some non-zero n-vectors u and v.

(a) If A is elementary, what condition on u and v ensures that A is non-singular?

(b) If A is elementary and non-singular, prove that A^{-1} is also elementary by showing that $A^{-1} = I - \sigma uv'$ for some scalar σ. What is the specific value of σ, in terms of u and v?

(c) Are elementary elimination matrices elementary? If so, what are u, v and σ in this case?

13. Prove the Sherman-Morrison formula

$$(A - uv')^{-1} = A^{-1} + A^{-1}u(1 - v'A^{-1}u)^{-1}v'A^{-1}.$$

Hint: Multiply both sides by $A - uv'$.

14. Prove the Woodbury formula

$$(A - UV')^{-1} = A^{-1} + A^{-1}U(I - V'A^{-1}U)^{-1}V'A^{-1}.$$

Hint: Multiply both sides by $A - UV'$.

15. Suppose the symmetric matrix

$$B = \begin{bmatrix} \alpha & a' \\ a & A \end{bmatrix}$$

of order $n + 1$ is positive definite.

(a) Show that the scalar α must be positive and that $n \times n$ matrix A must be positive definite.

(b) What is the Choleski factorization of B in terms of its constituent sub-matrices?

16. Suppose the symmetric matrix

$$B = \begin{bmatrix} A & a \\ a' & \alpha \end{bmatrix}$$

of order $n + 1$ is positive definite.

(a) Show that the scalar α must be positive and that $n \times n$ matrix A must be positive definite.

(b) What is the Choleski factorization of B in terms of its constituent sub-matrices?

17. Use Gaussian elimination without pivoting to solve the linear system

$$A(\epsilon)x(\epsilon) = b(\epsilon), \qquad A(\epsilon) = \begin{bmatrix} \epsilon & 1 \\ 1 & 1 \end{bmatrix}, \qquad b(\epsilon) = \begin{bmatrix} 1+\epsilon \\ 2 \end{bmatrix}$$

for $\epsilon = 10^{-2k}$, $k = 1, \ldots, 10$. I.e., perform one elimination step using the elementary elimination matrix

$$M(\epsilon) = \begin{bmatrix} 1 & \\ -1/\epsilon & 1 \end{bmatrix}$$

and then backsolve the resulting system $M(\epsilon)A(\epsilon)x = M(\epsilon)b(\epsilon)$.

Clearly, the exact solution $x(\epsilon)$ is $[1, 1]'$, independently of the value of ϵ. How does the accuracy of the computed solution behave as the value of ϵ decreases?

18. The 2-norm of a matrix A is defined as

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

How can $\| \cdot \|_2$ be expressed in terms of the singular values of A? Implement $\| \cdot \|_2$.

19. The condition number of a regular square matrix A is defined as $\kappa(A) = \|A\|\|A^{-1}\|$. If the 2-norm is used, how can κ be expressed in terms of the singular values of A? Implement κ.

20. Consider a linear system $Ax = b$ for a regular matrix A. We are interested in estimating the effect of changes in the rhs b on the solutions x. Show that the relative error in x does not exceed the relative error in b multiplied by the condition number of A, i.e.,

$$\frac{\|\Delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\Delta b\|}{\|b\|}$$

where Δb and Δx are the changes in b and x, respectively. (The bound is rather pessimistic.) (Hint: note that $\Delta x = A^{-1}\Delta b$ and $\|b\| = \|Ax\| \leq \|A\|\|x\|$.)

21. Consider the linear system

$$\begin{bmatrix} 1 & 1+\epsilon \\ 1-\epsilon & 1 \end{bmatrix} x = \begin{bmatrix} 1+\epsilon+\epsilon^2 \\ 1 \end{bmatrix}$$

where ϵ is a small parameter. The exact solution is obviously given by $x = [1, \epsilon]'$. Experiment with available methods for solving linear systems and ϵ small (especially near the square root of the machine precision). How does the accuracy obtained for each component compare with expectations based on the condition number of the matrix?

22. Suppose the $n \times n$ matrix A has the block upper triangular form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix}$$

where A_{11} is $k \times k$ and A_{22} is $(n-k) \times (n-k)$.

   (a) If λ is an eigenvalue of A_{11}, show that it is also an eigenvalue of A. (*Hint*: let u be the corresponding eigenvector if A_{11}, and determine an $(n-k)$-vector v such that $[u', v']'$ is an eigenvector of A with eigenvalue λ.)

   (b) If λ is an eigenvalue of A_{22} (but not of A_{11}), show that it is also an eigenvalue of A.

   (c) If λ is an eigenvalue of A with corresponding eigenvector $[u', v']'$ where u is a k-vector, show that λ is an eigenvalue of A_{11} with corresponding eigenvector u or an eigenvalue of A_{22} with corresponding eigenvector v.

   (d) Conclude that λ is an eigenvalue of A if and only if it is an eigenvalue of either A_{11} or A_{22}.

23. Let A be an $n \times n$ matrix of rank one so that $A = uv'$ for non-zero n-vectors u and v.

   (a) Show that $u'v$ is an eigenvalue of A.

   (b) What are the other eigenvalues of A?

   (c) What is the singular value decomposition of A?

   (d) Consider the power iteration $x_k = Ax_{k-1}$. How many iterations are required so that x_k exactly becomes an eigenvector corresponding to the dominant eigenvalue $u'v$ of A?

24. Show that for two n-vectors u and v, $\det(I + uv') = 1 + u'v$.

25. What are the eigenvalues of the Householder transformation

$$H = I - 2\frac{vv'}{v'v}$$

where v is a non-zero n-vector? What is the geometric interpretation of H?

26. A singular matrix must have a zero eigenvalue, but must a nearly singular matrix have a "small" eigenvalue? Consider a matrix of the form

$$\begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ & 1 & -1 & -1 & -1 \\ & & 1 & -1 & -1 \\ & & & 1 & -1 \\ & & & & 1 \end{bmatrix}$$

whose eigenvalues are obviously all ones. Compute the singular value decomposition of such a matrix for various dimensions. How does the ratio $\sigma_{\max}/\sigma_{\min}$ grow as the order of the matrix grows?

27. The matrix

$$C = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\gamma_0 \\ 1 & 0 & \cdots & 0 & -\gamma_1 \\ 0 & 1 & \cdots & 0 & -\gamma_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\gamma_{n-1} \end{bmatrix}$$

is called the *companion matrix* of the polynomial $p(z) = \gamma_0 + \gamma_1 z + \cdots + \gamma_{n-1} z^{n-1} + z^n$. Show that $(-1)^n p(z)$ is the characteristic polynomial $\det(C - zI)$ of its companion matrix.

Specifically, consider the polynomial

$$p(z) = 24 - 40z + 35z^2 - 13z^3 + z^4.$$

Compute all its roots by forming the companion matrix and determining its eigenvalues.

28. How can the singular value decomposition $A = UDV'$ of an $m \times n$ matrix A be used to obtain orthonormal bases for the range of A (the set $\{Ax : x \in \mathbb{R}^n\}$) and the null space of A (the set $\{x \in \mathbb{R}^n : Ax = 0\}$?

Implement two R functions for computing these bases. For this, you will need to determine which singular values are "numerically zero": a common criterion is to test whether singular values are less than $\sigma_1 \max(m, n)\epsilon$ for a given tolerance ϵ (e.g., `.Machine$double.eps`).

29. The "vec" operator stacks the columns of a matrix one underneath the other. I.e., if $A = [\alpha_{ij}]$, then

$$\mathrm{vec}(A) = (\alpha_{11}, \dots, \alpha_{m1}, \alpha_{12}, \dots, \alpha_{m2}, \dots, \alpha_{1n}, \dots, \alpha_{mn}).$$

Implement vec.

30. For symmetric matrices, e.g. the supradiagonal elements are redundant. The "vech" operator extracts the non-supradiagonal elements of an $n \times n$ matrix A as follows:

$$\mathrm{vech}(A) = (\alpha_{11}, \dots, \alpha_{n1}, \alpha_{22}, \dots, \alpha_{n2}, \dots, \alpha_{nn}).$$

Implement vech.

31. The *duplication matrix* D_n is defined as the unique matrix satisfying

$$D_n \mathrm{vech}(A) = \mathrm{vec}(A)$$

for all symmetric $n \times n$ matrices A, and can be used to recover a symmetric A from its non-redundant vech elements. Write a function returning D_n for given n.

32. Investigate the singular values of the duplication matrix D_n.

33. The *elimination matrix* L_n is the unique matrix satisfying

$$L_n \text{vec}(A) = \text{vech}(A)$$

for all $n \times n$ matrices A. Write a function returning L_n for given n.

34. What is the singular value decomposition of L_n?

35. Let A be an arbitrary $m \times n$ matrix. Clearly, vec(A) and vec(A') contain the same elements, but arranged in different order. Hence, there exists a unique $mn \times mn$ matrix K_{mn}, the *commutation matrix*, for which

$$K_{mn} \text{vec}(A) = \text{vec}(A').$$

Write a function returning K_{mn} for given m and n.

36. Investigate the singular values of the commutation matrix K_{mn}.

37. How can the generalized inverse (Moore-Penrose inverse) of a matrix A be expressed in terms of its singular value decomposition $A = UDV'$? Implement the Moore-Penrose inverse, providing a tolerance parameter `tol` controlling when singular values are regarded as zero.

38. Write an efficient function `wcptrace` computing trace($A'\text{diag}(w)A$) for given A and w.

39. The density of the multivariate normal distribution with mean m and regular covariance matrix V is given by

$$\det(2\pi V)^{-1/2} \exp(-(x-m)'V^{-1}(x-m)/2).$$

Write a computationally efficient function `dmvnorm` computing the density for a given matrix `x` containing the points as its rows, `m` and `V`. (E.g., try exploiting the potential of the eigendecomposition of V.)

40. This problem concerns the computation in `R` of the square root of a symmetric non-negative-definite square matrix.

Write a function, call it `msqrt`, with argument A which: Checks that A is a square matrix, is symmetric and diagonalizes the matrix by computing the eigenvalues and the matrix of the eigenvectors (hint: `eigen`). Return a symmetric matrix of the same size as A, with the same eigenvectors, the eigenvalue corresponding to a given eigenvector being the square root of the corresponding eigenvalues of A. The matrix returned is called the square root of the matrix A and will be denoted by $A^{1/2}$.

41. If X has a multivariate normal distribution with mean m_X and covariance matrix Σ_X, the linearly transformed $Y = AX + b$ has a multivariate normal distribution with mean $m_Y = Am_X + b$ and covariance matrix $\Sigma_Y = A\Sigma_X A'$. Use this result to obtain an algorithm for sampling from a multivariate normal distribution with parameters m and Σ based on independent, identically distributed samples from the (univariate) standard normal distribution (as obtained from `rnorm`), using (a) the eigendecomposition of Σ or (b) the Choleski decomposition of Σ. Write a function `rmvnorm` for generating n random points (collected in a matrix with n rows) from the multivariate normal distribution with mean m and covariance matrix Σ for given n, m and Σ.

42. Suppose X is a d-dimensional random vector with standardized margins (i.e., zero mean and unit variance) and an *equicorrelation* matrix, i.e, $\text{cor}(X_i, X_j) = \rho$ for $i \neq j$. Then $\text{cov}(X) = \rho 1_d 1_d' + (1-\rho)I_d$, where 1_d is a d-dimensional vector of ones. Assume that $\rho > 0$.

(a) Suppose that Y is a standardized scalar random variable independent of X. Write

$$F = \frac{\sqrt{\rho}}{1+\rho(d-1)} \sum_{j=1}^{d} X_j + \sqrt{\frac{1-\rho}{1+\rho(d-1)}} Y, \qquad \epsilon_i = X_i - \sqrt{\rho} F.$$

so that $X_i = \sqrt{\rho}F + \epsilon_i$. Show that $F, \epsilon_1, \ldots, \epsilon_d$ have zero means and are mutually uncorrelated with $\mathrm{var}(F) = 1$ and $\mathrm{var}(\epsilon_i) = 1 - \rho$.

(b) Suppose that X is Gaussian. Show that we have the one-dimensional factor model

$$X_i = \sqrt{\rho}F + \sqrt{1-\rho} Z_i,$$

where $F, Z_1, \ldots, Z_d$ are i.i.d. standard normal. Use this model to write a function for generating n random points in $\mathbb{R}^d$ (collected in a matrix with n rows) from a d-dimensional standard normal distribution with common correlation ρ, without using matrix decompositions.

43. A d-dimensional random vector X has a (multivariate) normal variance mixture distribution if

$$X =^d m + \sqrt{W} A Z,$$

where for some k

(a) $Z \sim N_k(0, I_k)$;

(b) W is a non-negative, scalar-valued random variable which is independent of Z;

(c) $A \in \mathbb{R}^{d \times k}$ and $m \in \mathbb{R}^d$ are a matrix and a vector of constants, respectively

(and $=^d$ denotes equality in distribution). Show that if W has finite expectation, then

$$\mathbb{E}(X) = m, \qquad \mathrm{cov}(X) = \mathbb{E}(W)\Sigma, \quad \Sigma = AA'.$$

Note that the distribution of X depends on A only via Σ.

If W has an inverse Gamma distribution with parameters $\nu/2$ and $\nu/2$, which is equivalent to saying that ν/W has a chi-squared distribution with ν degrees of freedom, then X has a *multivariate t distribution* $t_d(\nu, m, \Sigma)$ with parameters ν (degrees of freedom), m (mean) and Σ (dispersion). Show that if $\nu > 2$,

$$\mathbb{E}(X) = m, \qquad \mathrm{cov}(X) = \frac{\nu}{\nu - 2}\Sigma$$

(so that Σ is *not* the covariance matrix of X).

Write a function `rmvt` for generating n random points (collected in a matrix with n rows) from the multivariate t distribution with parameters ν, m and Σ. Test your implementation for some $d > 1$ for $\nu = 4$ and 13, $m = 0$ and Σ having diagonal entries one and off-diagonal entries $\rho = 0.3$, e.g., using Q-Q plots for the marginal distributions and comparing empirical and theoretical first and second moments.

(Note that if $d = k = 1$, $m = \mu$ and $A = \sigma > 0$, we obtain $X = \mu + \sigma\sqrt{W}Z$, i.e., a location-scale family for the "standard" variance mixture $\sqrt{W}Z$ (which is the standard univariate Student t distribution with ν degrees of freedom in the t case).)

44. Let $-1 < \rho < 1$.

First suppose that (X_1, X_2) has a bivariate standard normal distribution with correlation ρ. Show that

$$X_2 | X_1 = x \sim N(\rho x, 1 - \rho^2)$$

and use this to show that

$$\lim_{x \to -\infty} \mathbb{P}(X_2 \le x | X_1 = x) = 0$$

(i.e., extreme events occur independently).

Now suppose that (X_1, X_2) has a bivariate t distribution $t_2(\nu, 0, P)$ where P is a correlation matrix with off-diagonal element ρ. One can show that conditional on $X_1 = x$,

$$\sqrt{\frac{\nu+1}{\nu+x^2}} \frac{X_2 - \rho x}{\sqrt{1-\rho^2}}$$

has a standard univariate t distribution with $\nu+1$ degrees of freedom. Use this to determine $\lim_{x \to -\infty} \mathbb{P}(X_2 \le x | X_1 = x)$ (which is non-zero, so there is extreme tail dependence).

45. The goal of this problem is to prove rigorously a couple of useful formulae for normal random variables.

    (a) Show that if $Z \sim N(0,1)$, then for all β and functions f,

    $$\mathbb{E}(f(Z)e^{\beta Z}) = e^{\beta^2/2}\mathbb{E}(f(Z+\beta)),$$

    (provided that the integrals exist), and use this formula to recover the well known fact

    $$\mathbb{E}(e^X) = e^{\mu+\sigma^2/2},$$

    whenever $X \sim N(\mu, \sigma^2)$.

    (b) We now assume that X and Y are jointly-normal mean-zero random variables and that h is *any* function. Prove that:

    $$\mathbb{E}(e^X h(Y)) = \mathbb{E}(e^X)\mathbb{E}(h(Y + \mathrm{cov}(X,Y))).$$

$\boxed{\pi}$ 46. The purpose of this exercise is to investigate the minimally and maximally possible correlations $\rho_{\min}$ and $\rho_{\max}$, respectively, of log-normal random variables (i.e., random variables whose logarithms are normally distributed).

    Let (X_1, X_2) have a bivariate normal distribution with zero means, variances 1 and σ^2, and correlation ρ.

    (a) Construct a regularly spaced grid of 100 σ^2 values from 0.1 to 20, and a regularly spaced grid of 21 ρ values from -1 to 1. For each pair (σ^2, ρ), generate a sample of size $n = 500$ from the above bivariate normal, and compute the sample correlation of $\exp(X_1)$ and $\exp(X_2)$. Produce a scatterplot of the sample correlations against the σ^2 values used for obtaining them. What can you see?

    (b) Now prove theoretically that

    $$\rho_{\min}(\sigma) = \frac{e^{-\sigma}-1}{\sqrt{(e-1)(e^{\sigma^2}-1)}}, \qquad \rho_{\max}(\sigma) = \frac{e^{\sigma}-1}{\sqrt{(e-1)(e^{\sigma^2}-1)}},$$

    and that $\lim_{\sigma\to\infty} \rho_{\min}(\sigma) = \lim_{\sigma\to\infty} \rho_{\max}(\sigma) = 0$, and compare the exact and simulated bounds. (*Hint:* use the moment generating function of the multivariate normal distribution.)

47. Generate 1000 uniform pseudorandom variates using `runif()`, and assign them to vector `U`, using an initial seed of 19908.

    (a) Compute the average, variance and standard deviation of the values.

(b) Compare the results with the true mean, variance and standard deviations.

(c) Compute the proportion of the values of `U` that are less than 0.6, and compare with the probability that a uniform random variable U is less than 0.6.

48. Simulate 10 000 values of a Uniform$(0, 1)$ random variable U_1 using `runif()`, and another 10 000 values of a Uniform$(0, 1)$ random variable U_2. Assign these to `U1` and `U2`, respectively. Since the values in `U1` and `U2` are approximately independent, we can view U_1 and U_2 as independent Uniform$(0, 1)$ random variables.

    (a) Estimate $\mathbb{E}(U_1 + U_2)$. Compare with the true value, and compare with an estimate of $\mathbb{E}(U_1) + \mathbb{E}(U_2)$.

    (b) Estimate $\text{var}(U_1 + U_2)$ and $\text{var}(U_1) + \text{var}(U_2)$. Are they equal? Should the true values be equal?

    (c) Estimate $\mathbb{P}(U_1 + U_2 \leq 1.5)$.

    (d) Estimate $\mathbb{P}(\sqrt{U_1} + \sqrt{U_2} \leq 1.5)$.

49. Suppose U_1, U_2 and U_3 are independent uniform random variables on the interval $(0, 1)$. Use simulation to estimate the following quantities:

    (a) $\mathbb{E}(U_1 + U_2 + U_3)$.

    (b) $\text{var}(U_1 + U_2 + U_3)$ and $\text{var}(U_1) + \text{var}(U_2) + \text{var}(U_3)$.

    (c) $\mathbb{E}(\sqrt{U_1 + U_2 + U_3})$.

    (d) $\mathbb{P}(\sqrt{U_1} + \sqrt{U_2} + \sqrt{U_3} \geq 0.8)$.

50. Suppose a class of 100 writes a 20-question True-False test, and everyone in the class guesses at the answers.

    (a) Use simulation to estimate the average mark on the test as well as the standard deviation of the marks.

    (b) Estimate the proportion of students who would obtain a mark of 30% or higher.

51. Simulate 10 000 binomial pseudorandom numbers with parameters 20 and 0.3. Let X be a binomial$(20, 3)$ random variable. Use the simulated numbers to estimate the following:

    (a) $\mathbb{P}(X \leq 5)$

    (b) $\mathbb{P}(X = 5)$

    (c) $\mathbb{E}(X)$

    (d) $\text{var}(X)$

    (e) the 95th percentile of X (you may use the `quantile()` function)

    (f) the 99th percentile of X

    (g) the 99.9999th quantile of X.

    In each case, compare your estimates with the true values. What is required to estimate extreme quantities accurately?

52. Consider the following function which is designed to simulate binomial pseudorandom variates using the inversion method:

```
R> ranbin1 <- function(n, size, prob) {
+     cumbins <- pbinom(0 : (size - 1), size, prob)
+     singlenumber <- function() {
+         x <- runif(1)
+         sum(x > cumbins)
+     }
+     replicate(n, singlenumber())
+ }
```

(a) Study this function carefully and write documentation for it.

(b) Use `ranbin1()` to simulate vectors of length 1000, 10 000, and 100 000 from the binomial distribution with size parameter 10 and probability parameter 0.4. Use `system.time()` to compare the execution times for these simulations with the corresponding execution times when `rbinom()` is used.

53. The following function simulates binomial pseudorandom numbers by summing up the corresponding independent Bernoulli variates:

```
R> ranbin2 <- function(n, size, prob) {
+     singlenumber <- function(size, prob) {
+         x <- runif(size)
+         sum(x < prob)
+     }
+     replicate(n, singlenumber(size, prob))
+ }
```

(a) Study this function carefully and write documentation for it.

(b) Use `ranbin2()` to simulate vectors of length 10 000 from the binomial distribution with size parameters 10, 100, and 1000, and probability parameter 0.4. Use `system.time()` to compare the execution times for these simulations with the corresponding execution times when `rbinom()` is used, and compare timings with those for `ranbin1()`.

54. The generator for `ranbin2()` required `size` uniform pseudorandom numbers to be generated for each binomial number generated. The following generator is based on the same principle, but requires only one uniform pseudorandom number for each binomial number generated:

```
R> ranbin3 <- function(n, size, prob) {
+     singlenumber <- function(size, prob) {
+         k <- 0
+         U <- runif(1)
+         X <- numeric(size)
+         while(k < size) {
+             k <- k + 1
+             if(U <= prob) {
+                 X[k] <- 1
+                 U <- U / prob
+             } else {
+                 X[k] <- 0
+                 U <- (U - prob) / (1 - prob)
+             }
+         }
+         sum(X)
+         }
+     replicate(n, singlenumber(size, prob))
+ }
```

(a) Use the `ranbin3()` function to generate 100 pseudorandom numbers from the binomial$(20, 0.4)$ and binomial$(500, 0.7)$ distributions, respectively.

(b) What is the conditional distribution of U/p, given that $U < p$?

(c) What is the conditional distribution of $(U - p)/(1 - p)$, given that $U > p$?

(d) Use the answers to the above questions to provide documentation for `ranbin3()`.

55. Use random trials to determine the smallest number of persons required for the probability that two persons in a group have the same birthday to be greater than one half. Also justify your result analytically.

56. Use random sampling to determine the probability that the quadratic equation $ax^2+bx+c = 0$ has only real roots, if each of its coefficients a, b and c is randomly chosen from the interval $[-1, 1]$. Also justify your result analytically.

57. Let X be a random variable with cdf F and corresponding quantile function $F^{-1}(u) = \inf\{x : F(x) \geq u\}$. Show that for $x \in \mathbb{R}$ and $0 \leq u \leq 1$,

$$F^{-1}(u) \leq x \quad \Leftrightarrow \quad F(x) \geq u.$$

Conclude that

$$F(F^{-1}(u)) \geq u, \qquad F^{-1}(F(x)) \leq x$$

and characterize the cases where there is strict inequality.

58. Let X be a random variable with continuous cdf F. Show that the *probability integral transform* $F(X)$ is uniformly distributed on $[0, 1]$.

59. Let X be a discrete random variable which attains the values $x_1 < \cdots < x_n$ with probabilities $p_1, \ldots, p_n$. With F the cdf of X, what is the cdf of $F(X)$?

60. The Pareto(a, b) distribution has cdf

$$F(x) = 1 - (b/x)^a, \qquad x \geq b > 0, \quad a > 0.$$

Derive the probability inverse transformation $F^{-1}(U)$ and use the inverse transform method to simulate a random sample from the Pareto$(2, 2)$ distribution. Graph the density of the histogram of the sample with the Pareto$(2, 2)$ density superimposed for comparison.

61. The generalized Pareto distribution has cdf

$$F(x) = 1 - (1 + \xi(x - \mu)/\sigma)^{-1/\xi}$$

for in the support of this distribution, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. Establish first that for $\xi \to 0$, $F(x) \to 1 - \exp((-(x - \mu)/\sigma)$. Then show that the support is thus $[\mu, \infty)$ for $\xi \geq 0$ and $[\mu, \mu - \sigma/\xi]$ for $\xi < 0$.

62. The generalized Pareto distribution has cdf

$$F(x) = 1 - (1 + \xi(x - \mu)/\sigma)^{-1/\xi}$$

for $x \geq \mu$ and $x \leq \mu - \sigma/\xi$ when $\xi < 0$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. Use the inverse transform method to generate a random sample from the generalized Pareto distribution.

63. A discrete random variable X has probability mass function

x	0	1	2	3	4
$p(x)$	0.1	0.2	0.2	0.2	0.3

Use the inverse transform method to generate a random sample of size 1000 from the distribution of X. Construct the relative frequency table and compare the empirical with the theoretical probabilities. Repeat using function `sample()`.

64. Conduct a simulation experiment to check the reasonableness of the assertion that the distribution of the number of points from a rate 1.5 Poisson process which fall in the interval $[4, 5]$ is Poisson with a mean of 1.5. First, simulate a large number of realizations of the Poisson process on the interval $[0, 10]$. Then count the number of points in $[4, 5]$ for each realization. Compare this set of counts with simulated Poisson counts using a Q-Q plot.

65. A simple electronic device consists of two components which have failure times which may be modeled as independent exponential random variables. The first component has a mean time to failure of 3 months, and the second has a mean time to failure of 6 months. The electronic device will fail when either of the components fails. Use simulation to estimate the mean and the variance of the time to failure for the device. Re-do the calculations for the case where the device will fail only when both components fail.

66. A χ^2 random variable with n degrees of freedom has the same distribution as the sum of the squares of n independent standard normal random variables. Simulate a χ^2 random variable with 8 degrees of freedom, and estimate its mean and variance. (Compare with the theoretical values: 8 and 16.)

67. The following function returns normal pseudorandom numbers:

```
R> rannorm <- function(n, mean = 0, sd = 1){
+     singlenumber <- function() {
+         repeat {
+             U <- runif(1)
+             U2 <- sign(runif(1, min = -1)) # value is +/- 1.
+             Y <- rexp(1) * U2              # Y is a double exponental r.v.
+             if (U < dnorm(Y) / exp(-abs(Y))) break
+         }
+         Y
+     }
+     replicate(n, singlenumber()) * sd + mean
+ }
```

   (a) Use this method to generate a vector of 10 000 normal pseudorandom numbers with a mean of 8 and a standard deviation of 2.
   (b) Obtain a Q-Q plot to check the accuracy of this generator.
   (c) Use `curve()` to draw the graph of the standard normal density on the interval $[-4, 4]$. Use the `add = TRUE` parameter to overlay the exponential density on the same interval to verify that the rejection method has been implemented appropriately.

68. Consider the following two methods for simulating from the discrete distribution with values $0, 1, 2, 3, 4, 5$ which take respective probabilities $0.2, 0.3, 0.1, 0.15, 0.05, 0.2$.

    The first method is an inversion method:

```
R> probs <- c(0.2, 0.3, 0.1, 0.15, 0.05, 0.2)
R> randiscrete1 <- function(n, probs) {
+     cumprobs <- cumsum(probs)
+     singlenumber <- function() {
+         x <- runif(1)
+         sum(x > cumprobs)
+     }
+     replicate(n, singlenumber())
+ }
```

    The second method is a rejection method:

```
R> randiscrete2 <- function(n, probs) {
+     singlenumber <- function() {
+         repeat {
+             U <- runif(2,
+                        min = c(-0.5, 0),
+                        max = c(length(probs) - 0.5, max(probs)))
+             if(U[2] < probs[round(U[1]) + 1]) break
+         }
+         return(round(U[1]))
+     }
+     replicate(n, singlenumber())
+ }
```

Execute both functions using $n = 100$, 1000, and 10 000. Use `system.time()` to determine which method is faster.

69. Generate a random sample of size 1000 from a normal location mixture. The components of the mixture have $N(0,1)$ and $N(3,1)$ distributions with mixing probabilities p_1 and $p_2 = 1 - p_1$. Graph the histogram of the sample with density superimposed for $p_1 = 0.75$. Repeat with different values for p_1 and observe whether the empirical distribution of the mixture appears to be bimodal. Make a conjecture about the values of p_1 that produce bimodal mixtures.

70. Simulate a continuous Exponential-Gamma mixture. Suppose that the rate parameter Λ has Gamma(r, β) distribution and Y has Exp(Λ) distribution. That is, $Y|\Lambda = \lambda \sim f_Y(y|\lambda) = \lambda e^{-\lambda y}$. Generate 1000 random observations of this mixture with $r = 4$ and $\beta = 2$. It can be shown that the mixture has a Lomax ("Pareto" in actuarial science: a standard Pareto distribution shifted so that its support starts at 0) distribution with cdf

$$F(y) = 1 - \left(\frac{\beta}{\beta + y}\right)^r, \qquad y \geq 0.$$

Compare the empirical and theoretical (Lomax/Pareto) distributions by graphing the density histogram of the sample and superimposing the Lomax/Pareto density curve.

71. If (X_1, X_2) has a bivariate t distribution $t_2(\nu, m, \Sigma)$, one can show that the Kendall rank correlation $\rho_\tau(X_1, X_2)$ of X_1 and X_2 satisfies

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin(\rho),$$

where

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

is the *pseudo-correlation coefficient* and σ_{ij} denotes the (i, j) entry of Σ. (This relation more generally holds for pairs of random variables following a bivariate elliptical distribution.) Inverting this relation can be used for estimating ρ.

Generate $B = 3000$ samples of size $n = 90$ from a bivariate t distribution with $\nu = 3$ degrees of freedom and linear correlation $\rho = 0.5$, and estimate ρ (a) directly using sample correlations and (b) by inverting the sample Kendall's τ correlations.

Which method works better? How can this method be used to estimate the correlations of a d-variate t distribution with $d > 2$? Can you anticipate possible problems for this?

72. A *compound Poisson process* is a stochastic process $(X(t), t \geq 0$ which can represented as the random sum

$$X(t) = \sum_{i=1}^{N(t)} Y_i,$$

where $(N(t), t \geq 0)$ is a Poisson process and $Y_1, Y_2, \ldots$ are i.i.d. and independent of $(N(t))$. Write a program to simulate a compound Poisson(λ)–Gamma process. Estimate the mean and the variance of $X(10)$ for several choices of the parameters and compare with the theoretical values. *Hint:* show that $\mathbb{E}(X(t)) = \lambda t \mathbb{E}(Y_1)$ and $\mathrm{var}(X(t)) = \lambda t \mathbb{E}(Y_1^2)$.

73. Use Monte Carlo integration to estimate the following integrals, comparing with the exact answer if known.

$$\int_1^3 x^2 dx, \qquad \int_0^\pi \sin(x) dx, \qquad \int_0^\infty e^{-x} dx, \qquad \int_0^3 \sin(e^x) dx, \qquad \int_0^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

74. Compute a Monte Carlo estimate $\hat{\theta}$ of

$$\theta = \int_0^{0.5} e^{-x} dx$$

by sampling from Uniform$(0, 0.5)$, and estimate the variance of $\hat{\theta}$. Find another Monte Carlo estimator θ^* by sampling from the exponential distribution. Which of the variances (of $\hat{\theta}$ and θ^*) is smaller, and why?

75. Suppose a bank has a credit portfolio with 10 obligors with rating grade AA, 25 with A, and 96 rated BBB, respectively, with corresponding one-year probabilities of default (PDs) 0.0001, 0.0005, and 0.0025, respectively. Simulate default scenarios, assuming that the default occur independently.

$\boxed{\pi}$ 76. Consider a credit portfolio with n obligors. Suppose that obligor i defaults if its associated critical variable X_i falls below a default threshold d_i, and that the X_i are jointly normal with zero mean and unit variance and common asset correlation $\rho \geq 0$. (This is a simple Merton-type threshold model for portfolio credit risk.)

Take $n = 100$ and all d_i equal to the 2.5% quantile of the standard normal distribution. Let D_i be the default indicator for obligor i ($D_i = 1$ if obligor i defaults, $D_i = 0$ otherwise). Use simulation to determine the default correlation $\mathrm{cor}(D_i, D_j)$, $i \neq j$, as a function of the asset correlation ρ.

Can you determine the exact values of the default correlations?

(*Hint:* The X_i can be simulated via $X_i = \sqrt{\rho} Z_0 + \sqrt{1-\rho} Z_i$, where $Z_0, \ldots, Z_n$ are i.i.d. standard normal.)

$\boxed{\pi}$ 77. Consider a credit portfolio with n obligors. Given a common (latent) factor Ψ, defaults occur independently with probability $Q = p(\Psi)$. Let D_i be the default indicator for obligor i ($D_i = 1$ if obligor i defaults, $D_i = 0$ otherwise), and write $M = D_1 + \cdots + D_n$ for the total number of defaults.

Suppose that Q has a Beta distribution with parameters a and b.

(a) Take $n = 100$, $a = 1$ and $b = 9$, and use simulation to determine the default probabilities $\mathbb{E}(D_i)$, the default correlations $\mathrm{cor}(D_i, D_j)$, $i \neq j$, and the distribution of M.

(b) Obtain exact expressions for the default probabilities and correlations, and the distribution of M. How well does simulation work?

(The distribution of M is called the beta-binomial distribution.)

$\boxed{\pi}$ 78. The following model has been used for the study of contagion. Suppose that there are N persons some of whom are sick with influenza. The following assumptions are made:

- when a sick person meets a healthy one, the chance is α that the latter will be infected
- all encounters are between two persons
- all possible encounters in pairs are equally likely
- one such encounter occurs in every unit of time.

(a) Write a function which simulates this model for various values of N (say, 10 000) and α (say, between 0.001 and 0.1). Monitor the history of this process, assuming that one individual is infected at the beginning.

(b) Suppose that initially only one individual is infected. What is the expected length of time until 1000 people are infected?

(c) Now add the assumption that each infected person has chance β, say 0.01, of recovering at each time unit. Monitor several histories of this new process, and compare them with the histories of the old process.

(d) Re-do with the assumption that the time between encounters is an exponential random variable with a mean of 5 minutes.

(e) Re-do assuming that the time between encounters is the absolute value of a normal random variable with a mean of 5 minutes and a standard deviation of 1 minute.

$\boxed{\pi}$ 79. Simulate the following simple model of auto insurance claims:

- Claims arise according to a Poisson process at a rate of 100 per year.
- Each claim is random in size following a Gamma distribution with shape and rate parameters both equal to 2. This distribution has a mean of 1 and a variance of 1/2. Claims must be paid by the insurance company as soon as they arise.
- The insurance company earns premiums at a rate of 105 per year, spread evenly over the year (i.e., at time t measured in years, the total premium received is $105t$).

Write R code to do the following:

(a) Simulate the times and amounts of all the claims that would occur in one year. Draw the graph of the total number of money that the insurance company would have through the year, starting from zero: it should increase smoothly with the premiums, and drop at each claim time.

(b) Repeat the simulation 1000 times, and estimate the following quantities:

   i. The expected minimum amount of money that the insurance company would have.
   ii. The expected final amount of money that the insurance company would have.

$\boxed{\pi}$ 80. An insurance company has an initial risk reserve of $x = 5$ monetary units (MUs). Premiums flow in at a constant rate of $p = 1$ MU per year. Claims arrive according to a Poisson process with rate $\lambda = 2$, are independent identically exponentially distributed with rate $\delta = 2.25$, and must be paid once they arise. Thus, if $N(t)$ is the number of claims up to time t and X_i is the size of claim i, the risk reserve at time t is $R(t) = x + pt - \sum_{i:i\leq N(t)} X_i$. Use simulation to determine the *ruin probability*

$$\psi(x) = \mathbb{P}\left(R(t) < 0 \text{ for some } t < 0\right).$$

(In this simple case, the ruin probability can be determined analytically: can you find the corresponding expression?)

$\boxed{\pi}$ 81. For the German credit data, explore the relations between personal status ('Status_and_sex') and Age, Amount, Purpose and quality ('Class'). Which patterns can you find?

$\boxed{\pi}$ 82. For the German credit data, explore the relations between credit history ('History'), employment level ('Job') and duration ('Employment_since'), and credit amount ('Amount') and quality ('Class'). Which patterns can you find?

83. The R standard data set `islands` is a vector containing the areas of 48 land masses.

    (a) Plot a histogram of these data.

    (b) Are there advantages to taking logarithms of the areas before plotting the histogram?

    (c) Compare the histograms that result when using breaks based on Sturges' and Scott's rules. Make this comparison on the log-scale and the original scale.

    (d) Construct a boxplot for these data on the log-scale as well as on the original scale.

    (e) Construct a dot-chart of the areas. Is a log transformation needed here?

    (f) Which form of graphic do you think is most appropriate for displaying these data?

84. For the `islands` data set, try out the following code.

```
R> hist(log(islands, 10), breaks = "Scott", axes = FALSE,
+       xlab = "Area", main = "Histogram of Island Areas")
R> axis(1, at = 1 : 5, labels = 10 ^ (1 : 5))
R> axis(2)
R> box()
```

    (a) Explain what is happening at each step of the above code.

    (b) Add a subtitle to the plot such as "Base-10 Log-Scale".

    (c) Modify the code to incorporate the use of the Sturges rule in place of the Scott rule. In this case, you will need to use the `round()` function to ensure that excessive numbers of digits are not used in the axis labels.

85. The R standard data set `stackloss` is a data frame with 21 observations on four variables taken at a factory where ammonia is converted to nitric acid. The first three variables are `Air.Flow`, `Water.Temp` and `Acid.Conc`. The fourth variable is `stack.loss`, which measures the amount of ammonia that escapes before being absorbed. (Read the help file for more information about this data frame.)

    (a) Use scatterplots to explore possible relationships between acid concentration, water temperature, and air flow and the amount of ammonia which escapes. Do these relationships appear to be linear or nonlinear?

    (b) Use the `pairs()` function to obtain all pairwise scatterplots among the four variables. Identify pairs of variables where the might be linear or nonlinear relationships.

86. The R standard data set `pressure` is a data frame with two variables, `temperature` and `pressure`.

    (a) Construct a scatterplot with `pressure` on the vertical axis and `temperature` on the horizontal one. Are the variables related linearly or non-linearly?

    (b) The graph of the following function passes through the plotted points reasonably well:

    $$y = (0.168 + 0.007x)^{20/3}$$

    The differences between the pressure values predicted by the curve and the observed pressure values are called *residuals*, and can be computed via

```
R> residuals <-
+     with(pressure,
+          pressure - (0.168 + 0.007 * temperature)^(20/3))
```

Construct a normal Q-Q plot of these residuals and decide whether they are normally distributed or whether they follow a skewed distribution.

(c) Now, apply the power transformation $y^{3/20}$ to the pressure data values. Plot these transformed values against temperature. Is a linear or nonlinear relationship evident now?

(d) Calculate residuals for the difference between transformed pressure values and those predicted by the straight line. Obtain a normal Q-Q plot, and decide whether the residuals follow a normal distribution or not.

87. Consider the `pressure` data set.

(a) Plot `pressure` against `temperature`, and use the following command to pass a curve through these data:

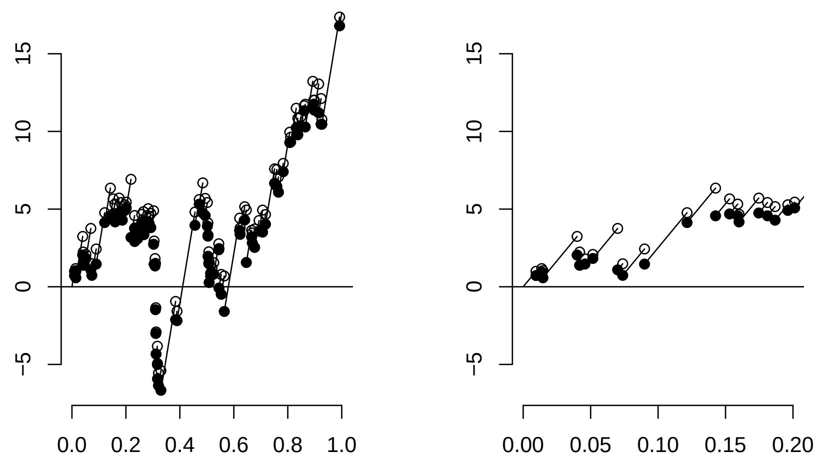
```
R> curve((0.168 + 0.007 * x)^(20/3), from = 0, to = 400, add = TRUE)
```

(b) Now, apply the power transformation $y^{3/20}$ to the pressure data values. Plot these transformed values against temperature. Is a linear or nonlinear relationship evident now? Use the `abline()` function to pass a straight line through the points. (See the previous part of the exercise to obtain an intercept and a slope for this line.)

(c) Add a suitable title to the graph.

(d) Re-do the above plots, but use the `mfrow` mechanism to display them in a 2×1 layout on the graphics page. Repeat once again using a 1×2 layout.

88. Write a function to "correctly" graph the amount of money that the insurance company would have in the auto insurance simulation model in the following style:



$\boxed{\pi}$ 89. How normal are stock returns?

Consider the daily log returns of (at least 20) or the 30 stocks in the Dow Jones Industrial Average for the past 10 years. Use **quantmod** to obtain these returns and compute sample skewness and kurtosis coefficients and the p-value of the Jarque-Bera test for normality (use `jarque.bera.test()` from package **tseries**). What do you find?

How do your results change when instead considering weekly, monthly or quarterly returns? (*Hint:* see `?quantmod::periodReturn` for conveniently obtaining these returns.)

$\boxed{\pi}$ 90. This problem deals with the analysis of the daily S&P 500 closing values.

(a) Load the data using **quantmod**.

(b) Create a vector DSPRET containing the daily raw returns. Compute the mean and the variance of this daily raw return vector, produce its eda.shape plot, and discuss the features of this plot which you find remarkable.

(c) Fit a Generalized Pareto Distribution (GPD for short) to the daily raw returns, give detailed plots of the fit in the two tails, and discuss your results.

(d) Generate a sample of size 10000 from the GPD fitted above. Call this sample SDSPRET, produce a Q-Q plot of DSPRET against SDSPRET, and comment.

(e) Compute the VaR (expressed in units of the current price) for a horizon of one day, at the level $\alpha = 0.01$ in each of the following cases:

- assuming that the daily raw return is normally distributed;
- using the GPD distribution fitted to the data in the previous question;
- compare the results (think of a portfolio containing a *very large* number of contracts).

$\boxed{\pi}$ 91. Redo the above problem after replacing the vector DSP with a vector SDSP containing only the first 6000 entries of DSP. Compare the results, and especially the VaRs. Explain the differences.

$\boxed{\pi}$ 92. For $0 < \alpha < 1$, the expected shortfall ES_α of a normal distribution with mean μ and variance σ^2 is

$$\mathrm{ES}_\alpha = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha},$$

where ϕ and Φ, respectively, are the density and distribution function of the standard normal distribution. For a univariate t distribution with ν degrees of freedom and location and scale parameters μ and σ one can show that

$$\mathrm{ES}_\alpha = \mu + \sigma \frac{f_{t_\nu}(F_{t_\nu}^{-1}(\alpha))}{1-\alpha} \frac{\nu + (F_{t_\nu}^{-1}(\alpha))^2}{\nu - 1},$$

where f_{t_ν} and F_{t_ν}, respectively, are the density and distribution function of the standard univariate t distribution with ν degrees of freedom.

Use **quantmod** to obtain the daily log returns of the stock prices of Boeing Co (BA), Walt Disney Co (DIS), General Electric Co (GE), and Microsoft Corp (MSFT) from 1993 to 2000. Fit normal distributions (by using the sample means and variances) and t location-scale distributions (see `?MASS::fitdistr` for how to do this) to these log returns. Estimate the value-at-risk $\mathrm{VaR}_{0.99}$ and the expected shortfall $\mathrm{ES}_{0.975}$ of the daily losses (a) directly (using the sample quantities), (b) using the fitted normal model and (c) using the fitted t model. Interpret your results.

93. What is the relation between expected shortfall and value at risk for large α? First try to numerically determine the limit of the *shortfall-to-quantile ratio* $\mathrm{ES}_\alpha/\mathrm{VaR}_\alpha$ as $\alpha \to 1-$ for losses from a normal distribution, or a Student t distribution with shape parameter $\nu > 1$. Then prove that the limit is 1 for the normal, and $\nu/(\nu-1)$ for Student's t.

(*Hint:* For the normal, use that for $x \to \infty$, $\Phi(x) \approx 1 - \phi(x)/x$. For Student's t, note that $f_{t_\nu}(x) \approx c(\nu)/x^{\nu+1}$ for $x \to \infty$, and use this to derive an asymptotic approximation for $F_{t_\nu}^{-1}(\alpha)$ as $\alpha \to 1-$).

94. Determine VaR_α and ES_α for the generalized Pareto distribution with location parameter μ, scale parameter σ and shape parameter $\xi < 1$. What is the limit of the *shortfall-to-quantile ratio* $\mathrm{ES}_\alpha/\mathrm{VaR}_\alpha$ as $\alpha \to 1-$?

$\boxed{\pi}$ 95. Use **quantmod** to obtain the daily log returns of the stock prices of International Business Machines Co (IBM), The Coca-Cola Co (KO), McDonald's Corp (MCD) and Nike Inc (NKE) from 1993 to 2000. Fit (symmetric) t location-scale distributions see (`?MASS::fitdistr` for how to do this) and 2-sided Generalized Pareto distributions to these log returns, and discuss the implied tail behavior. (Note that the tail probabilities of the t and GPD distributions behave, respectively, like $x^{-\nu}$ and $x^{-1/\xi}$.)

96. Let $U_1, \dots, U_d$ be i.i.d. $U[0,1]$ random variables. Determine the copulas of (a) $(U_1, \dots, U_d)$ (independence copula), (b) $(U_1, \dots, U_1)$ (U_1 repeated d times, comonotonicity copula), and (c) $(U_1, 1-U_1)$ (countermonotonicity copula).

97. Show that for every copula $C(u_1, \dots, u_d)$ we have

$$\max\left(\sum_{i=1}^{d} u_i + 1 - d, 0\right) \le C(u_1, \dots, u_d) \le \min(u_1, \dots, u_d).$$

98. Consider a bivariate Bernoulli random variable (X_1, X_2) with joint distribution

$$\mathbb{P}(X_1 = 0, X_2 = 0) = 1/8, \quad \mathbb{P}(X_1 = 1, X_2 = 1) = 3/8$$

and

$$\mathbb{P}(X_1 = 0, X_2 = 1) = \mathbb{P}(X_1 = 1, X_2 = 0) = 2/8.$$

Determine all possible copulas for (X_1, X_2) (i.e., all C such that for all x_1, x_2 we have $\mathbb{P}(X_1 \le x_1, X_2 \le x_2) = C(\mathbb{P}(X_1 \le x_1), \mathbb{P}(X_2 \le x_2)))$.

99. Let X and Y be continuous random variables with cdfs F_X and F_Y, respectively, and copula C. Prove that

(a) $\mathbb{P}(\max(X, Y) \le t) = C(F_X(t), F_Y(t))$,

(b) $\mathbb{P}(\min(X, Y) \le t) = F_X(t) + F_Y(t) - C(F_X(t), F_Y(t))$.

100. Show that if X is a random variable with distribution symmetric about 0 and finite second moments, then X and $|X|$ are uncorrelated.

101. This elementary exercise is intended to give an example showing that lack of correlation does not necessarily mean independence. Let us assume that $X \sim N(0,1)$ and let us define the random variable Y by $Y = (|X| - \sqrt{2/\pi})/\sqrt{1 - 2/\pi}$.

(a) Compute $\mathbb{E}(|X|)$.

(b) Show that Y has mean zero, variance 1, and that it is uncorrelated with X.

102. The purpose of this problem is to show that lack of correlation does not imply independence, even when the two random variables are Gaussian. We assume that X, ϵ_1 and ϵ_2 are independent random variables, that $X \sim N(0,1)$, and that $\mathbb{P}(\epsilon_i = -1) = \mathbb{P}(\epsilon_i = +1) = 1/2$ for $i = 1, 2$. We define the random variable X_1 and X_2 by $X_1 = \epsilon_1 X$, and $X_2 = \epsilon_2 X$.

(a) Prove that $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$ and that $\rho(X_1, X_2) = 0$.

(b) Show that X_1 and X_2 are not independent.

(c) Determine the copula of X_1 and X_2.

103. Let the random variables X_1 and X_2 represent the profits and losses on two portfolios, and suppose we are told that both have standard normal distributions and are uncorrelated. One possible model for this is taking X_1 and X_2 as independent (Model A). Another possibility is taking $(X_1, X_2) = (Z, \epsilon Z)$ where Z is standard normal and ϵ takes the values ± 1 with probabilty $1/2$ and is independent of Z (Model B). This roughly corresponds to a situation where with equal probability, profits and losses are perfectly comonotonic (either gain or lose the same in both portfolios) or perfectly countermonotonic (gain in one portfolio and lose the same in the other).

Verify that for model B we indeed have standard normal margins and zero correlation, and that the copula is

$$C_B(x_1, x_2) = \frac{1}{2}\max(u_1 + u_2 - 1, 0) + \frac{1}{2}\min(u_1, u_2).$$

Compute and compare the VaRs for the aggregate profits and losses $X_1 + X_2$ under model A and model B.

104. A d-dimensional copula is a distribution function on $[0,1]^d$ with standard uniform marginal distributions. I.e., if $(U_1, \dots, U_d)$ is a random vector such that $U_i \sim U[0,1]$, then

$$C(u_1, \dots, u_d) := \mathbb{P}(U_1 \le u_1, \dots, U_d \le u_d)$$

is the corresponding copula. Note that $U_1, \dots, U_d$ do not need to be independent!

Assume that $F_1, \dots, F_d$ are arbitrary continuous and increasing distribution functions, let $F_1^{-1}, \dots, F_d^{-1}$ be their corresponding quantile functions, and let $Z := (F_1^{-1}(U_1) \dots, F_d^{-1}(U_d))$. Show that

$$\mathbb{P}(Z_1 \le z_1, \dots, Z_d \le z_d) = C(F_1(z_1), \dots, F_d(z_d)).$$

Note. The converse statement is also true: for a given arbitrary random vector Z with marginal distributions $F_1, \dots, F_d$ we can find a copula C such that

$$\mathbb{P}(Z_1 \le z_1, \dots, Z_d \le z_d) = C(F_1(z_1), \dots, F_d(z_d)).$$

This result is known as *Sklar's Theorem.*

105. Let $\psi : [0, \infty) \to [0,1]$ be continuous and non-increasing with $\psi(0) = 1$ and $\lim_{t\to\infty} \psi(t) = 0$. For $0 \le u \le 1$, let $\psi^{-1}(u) = \inf\{t : \psi(t) = u\}$ (where the inf is always attained for $u > 0$). Show that if ψ is also convex, then

$$C(u, v) = \psi(\psi^{-1}(u) + \psi^{-1}(v))$$

is a copula, the so-called *Archimedean copula* with generator ψ. (One can show that convexity is also necessary for $\psi(\psi^{-1}(u) + \psi^{-1}(v))$ to be a copula.)

106. The Gumbel family of copulas has

$$C_\theta^{\mathrm{Gu}}(u, v) = \exp(-((-\log(u))^\theta + (-\log(v))^\theta)^{1/\theta})$$

where $\theta \ge 1$. Show that this is an Archimedean copula, and determine its generator. What are the limits for $\theta \to 1+$ and $\theta \to \infty$?

107. Copulas are used to model dependence structures of multivariate random variables. Of particular importance is the *Gaussian copula*, which is the corresponding copula to a multivariate normal distribution with mean zero and correlation matrix P.

In the spirit of Sklar's theorem we use the following algorithm to simulate from the copula:

(a) Generate $(Z_1, \dots, Z_d) \sim N_d(0, P)$ using `rmvnorm`

(b) Return $U = (\Phi(Z_1), \ldots, \Phi(Z_d))$, where Φ denotes the cdf of the standard normal distribution. The random vector U is now distributed according to the Gaussian copula with parameter P.

How does P influence the dependence structure of U? Generate 2000 samples each for for

- $\mathrm{cov}(Z_1, Z_2) = 0$
- $\mathrm{cov}(Z_1, Z_2) = -0.8$
- $\mathrm{cov}(Z_1, Z_2) = 0.8$

(note that $\mathrm{var}(Z_1) = \mathrm{var}(Z_2) = 1$). For each case, plot your results as a scatter plot. What can you see? What happens if $\mathrm{cov}(Z_1, Z_2) \to \pm 1$?

108. The Student t *copula* is the copula of a multivariate t distribution $t_d(\nu, 0, P)$ where P is a correlation matrix.

In the spirit of Sklar's theorem we can use the following algorithm to simulate from a t copula:

(a) Generate $(X_1, \ldots, X_d) \sim t_d(\nu, 0, P)$

(b) Return $U = (F_{t_\nu}(X_1), \ldots, F_{t_\nu}(X_d))$, where F_{t_ν} is the distribution function of the standard univariate t distribution with ν degrees of freedom.

Implement a function which generates n random points (collected in a matrix with n rows) from the multivariate t copula with given parameters ν and P.

Investigate how P influences the dependence structure of U by inspecting scatterplots of generated samples with $\nu = 3$ or 10, $d = 2$ and off-diagonal correlations ρ of -0.7, 0 and 0.7.

$\boxed{\pi}$ 109. This problem is based on the data matrix `UTILITIES`. Each row corresponds to a given day. The first column gives the log of the weekly return on an index based on Southern Electric stock value and capitalization (we'll call that variable X), and the second column gives, on the same day, the same quantity for Duke Energy (we'll call that variable Y), another large utility company.

(a) Compute the means and the standard deviations of X and Y, and compute their correlation coefficients.

(b) We first assume that X and Y are samples from a jointly Gaussian distribution with parameters computed in question 1. Compute the q-percentile with $q = 2\%$ of the variables $X + Y$ and $X - Y$.

(c) Fit a generalized Pareto distribution (GPD) to X and Y separately, and fit a copula of the Gumbel family to the empirical copula of the data.

(d) Generate a sample of size N (where N is the number of rows of the data matrix) from the joint distribution estimated in question 3. Use this sample to compute the same statistics as in question 1 (i.e., means and standard deviations of the columns, as well as their correlation coefficients), and compute the results to the numerical values obtained in question 1. Compute, still for this simulated sample, the two percentiles considered in question 2, compare the results, and comment.

$\boxed{\pi}$ 110. An alternative to **quantmod** for loading limited free financial data is offered by the package **Quandl**. See `https://cran.r-project.org/package=Quandl/README.html` for basic information about using this package to access Quandl data from R. For this exercise install the package **Quandl** from CRAN and load the daily prices of West Texas Intermediate crude oil (WTI) and Brent Crude Oil (Brent) from 2004-01-01 to 2014-01-01. The codes to use are `"FRED/DCOILWTICO"` and `"FRED/DCOILBRENTEU"`, respectively. Restrict the data to the days where prices for both crude oils are available, i.e., remove days where at least one has an `NA` entry. Compute log-returns for the remaining data. These log-returns should be considered in the sequel.

a) Compute the means for the WTI and Brent sample as well as their joint covariance matrix. Simulate a sample with the same size as the data from a bivariate normal distribution with the corresponding parameters. Compare scatter-plots of the actual data and the simulated data, compute Q-Q plots of actual data versus simulated data and comment.

b) To account for the heavier tails fit a multivariate t-distribution with location and scale parameter to WTI and Brent. This multivariate distribution is defined as follows. We start with an $\mathbb{R}^k$-valued random vector T whose components are independent and follow a t-distribution with ν degrees of freedom. With a vector $\mu \in \mathbb{R}^k$ and a matrix $\Sigma \in \mathbb{R}^{k\times k}$ with Choleski decomposition $\Sigma = LL'$ the multivariate t-distribution with ν degrees of freedom is defined as the distribution of X which is given by

$$X = \mu + LT.$$

The expectation and the covariance matrix are then ($\nu > 2$)

$$\begin{aligned}\mathbb{E}(X) &= \mu \\ \operatorname{cov}(X) &= L\operatorname{cov}(T)L' = \frac{\nu}{\nu-2}LL' = \frac{\nu}{\nu-2}\Sigma.\end{aligned} \tag{1}$$

Write a function which generates a sample of arbitrary size from a multivariate t-distribution with the given parameters μ, Σ and ν. Then choose parameters for a bivariate t-distribution with 4 degrees of freedom such that the resulting moments in (1) match the empirical moments. Simulate a sample with the same size as the data from this bivariate t-distribution with the corresponding parameters. Compare scatter-plots of the actual data and the simulated data and/or analyze Q-Q plots and comment.

c) Now match one-dimensional t-distributions with 4 degrees of freedom and location and scale parameter to WTI and Brent separately. Use these marginal distributions to fit a copula of the Gumbel family to the empirical copula of the data. Simulate a sample from this distribution and compare the scatter-plots again.

d) Assume now that an investor holds a portfolio with a value of USD 10 million. Fifty percent are invested in WTI and the remaining fifty percent are invested in Brent. Using simulation estimate the 1 percent quantile of the discrete daily return of this portfolio for all the distributions of log-returns considered (use `quantile`). Use these estimated quantiles to estimate the Value at Risk of the daily return. Compare the different estimates you get.

$\boxed{\pi}$ 111. The (in)famous Li model is/was a simple dynamic credit risk model used by practitioners to price basket credit derivatives. Given a set of n obligors, let τ_i by the time to default of obligor i. In the Li model, the τ_i are assumed to have exponential distributions with rates λ_i, with dependence given by a Gauss copula with suitable correlation matrix P.

Take $n = 125$, $\lambda_i \equiv 0.3$ and an equicorrelation copula (i.e., all off-diagonal entries of P are equal to the same ρ) with $\rho = 0.7$. Simulate the distribution of the number of defaults in one year, and in particular estimate the probability of no default in one year, as accurately as possible.