# Tinbergen Institute Statistics
# Exam questions

---

1. Let $U$ be a random variable that has a uniform distribution on $[0, 1]$. It is known that $\mathbb{E}\, U = \frac{1}{2}$ and that $\mathbb{V}\mathrm{ar}\, U = \frac{1}{12}$. Define other random variables $X$ and $Y$ by $X = a + (b - a)U$ for $a < b$ and $Y = -\theta \log U$ for some $\theta > 0$.

   (a) Use the transformation rule for densities to show that $X$ has a uniform distribution on $[a, b]$.

   (b) Compute $\mathbb{E}\, X$ and $\mathbb{V}\mathrm{ar}\, X$ from the definition of $X$.

   (c) Compute for each $y > 0$ the probability $\mathbb{P}(Y > y)$. What is the density of $Y$?

   (d) If $g(x) = x \log x - x$ (for $x > 0$), then $g'(x) = \log x$. Use this to compute $\mathbb{E}\, Y$.

2. Let $X_1, \ldots, X_n$ be a sample from a Poisson($\lambda$) distribution (so they are independent Poisson($\lambda$) distributed random variables).

   (a) Give the formula for $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$ (with $x_i$ nonnegative integers).

   (b) Show that $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the maximum likelihood estimator of $\lambda$.

   (c) Show that the Fisher information $I(\lambda)$ is equal to $1/\lambda$.

   (d) Use the Cramér-Rao bound to show that $\overline{X}$ has minimum variance among all unbiased estimators of $\lambda$.

   (e) Give a consistent estimator of $I(\lambda)$.

3. Let $U$ have a $\chi^2_m$ distribution and $V$ a $\chi^2_n$ distribution and assume that the random variables $U$ and $V$ are independent. The random variables $X_1, \ldots, X_n$ form a sample from the $N(0, \sigma^2)$ distribution, where $\sigma^2$ is unknown.

   (a) Use the definition of the $\chi^2$ distributions to show that $U + V$ has a $\chi^2_{m+n}$ distribution.

   (b) Show that the maximum likelihood estimator of $\sigma^2$ (call it $s^2$) is given by $\frac{1}{n} \sum_{i=1}^{n} X_i^2$.

   (c) Give a $(1 - \alpha)$-confidence interval for $\sigma^2$ based on $\sum_{i=1}^{n} X_i^2$.

(d) Suppose that one is interested in testing the null-hypothesis that the $X_i$ form a sample from the standard normal distribution against the alternative that the sample is from another normal distribution with zero expectation. Formulate this problem as a parameter testing problem.

(e) Suppose that one wants to perform a test based on $s^2$ for the above problem with significance level $\alpha = 0.01$. If $s^2 = 1.24$ is observed for $n = 40$ observations, will the null-hypothesis be rejected?

4. Consider the multivariate regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where the design matrix $X$ is of size $n \times p$, and where the elements $e_i$ of the vector $\mathbf{e}$ are independent random variables with $\mathbb{E}\, e_i = 0$ and $\mathbb{V}\mathrm{ar}\, e_i = \sigma^2$. The least squares estimator of $\beta$ is given by $\hat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\top\mathbf{Y}$. Let $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\top$ and $\mathbf{Q} = \mathbf{I_n} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\top$ ($\mathbf{I}_n$ is the $n$-dimensional identity matrix).

   (a) Let $\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$ and the residual vector $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}}$. Show that $\hat{\mathbf{Y}} = \mathbf{Pe} + \mathbf{X}\beta$, that $\hat{\mathbf{e}} = \mathbf{Qe}$ and that $\mathbf{PQ}$ is the zero matrix.

   (b) Show that $\mathbb{C}\mathrm{ov}(\hat{\mathbf{Y}}, \hat{\mathbf{e}}) = \mathbb{E}\, \hat{\mathbf{Y}}\hat{\mathbf{e}}^\top = \mathbf{0}$.

   (c) Suppose that one has an additional (row) vector of design variables $x_{n+1}$. The corresponding response variable $Y_{n+1}$ is then predicted by $\hat{Y}_{n+1} = x_{n+1}\hat{\beta}$. What is the expectation $\mathbb{E}\, \hat{Y}_{n+1}$ of $\hat{Y}_{n+1}$ and what the variance?

   (d) Suppose that we also know that the $e_i$ are $N(0, \sigma^2)$ distributed random variables with *known* $\sigma^2$. Construct a $(1 - \alpha)$-confidence interval for $x_{n+1}\beta$.

5. Let $Z$ be a random variable that has the standard normal distribution. By $\Phi$ we denote the distribution function of $Z$. Define another random variable $X$ by $X = Z^2$.

   (a) Show that the distribution function of $X$, $F$ say, is given by $F(x) = 2\Phi(\sqrt{x}) - 1$, for $x > 0$. What is $F(x)$ for $x \leq 0$?

   (b) Show that the density $f$ of $X$ is given by $f(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}x}$, for $x > 0$. Can we use the transformation rule to compute the density of $X$ directly from that of $Z$?

   (c) Let $Y = aX + b$, where $a$ and $b$ are constants. For which values of $a$ and $b$ do we have $\mathbb{E}\, Y = 0$ and $\mathbb{V}\mathrm{ar}\, Y = 1$? (You may use that $\mathbb{V}\mathrm{ar}\, X = 2$.)

6. Let $X_1, \dots, X_n$ be a sample from an exponential distribution with parameter $\lambda$ (so they are independent random variables with the same exponential distribution).

   (a) Give the formula for $f(x_1, \dots, x_n)$ (with the $x_i$ nonnegative real numbers), where $f$ is the joint probability density function of $(X_1, \dots, X_n)$.

   (b) Show that with $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ the maximum likelihood estimator of $\lambda$ is given by $1/\overline{X}$.

   (c) Show that the Fisher information $I(\lambda)$ is equal to $1/\lambda^2$.

   (d) Characterize the limit distribution of the maximum likelihood estimator of $\lambda$.

7. The random variables $X_1, \dots, X_{2n}$ form a sample of size $2n$ from the $N(0, \sigma^2)$ distribution, where $\sigma^2$ is *unknown.* Let $U$ be a random variable that is independent of the $X_i$ and that has a $\chi^2_n$ distribution. Moreover, we *observe* the $2n$ random variables $Y_i$ that are defined by $Y_i = X_i + \mu$, for some unknown number $\mu$.

   (a) What is the distribution of $\frac{1}{\sigma\sqrt{2}}(X_1 + X_2)$?

   (b) Show that $\frac{X_1+X_2}{\sigma\sqrt{2U/n}}$ has a $t_n$ distribution.

   (c) What is the distribution of each of the $Y_i$? Are the $Y_i$ independent?

   (d) Show that the maximum likelihood estimator (based on the observed random variables) of $\mu$ (call it $\hat{\mu}$) is given by $\frac{1}{2n}\sum_{i=1}^{2n} Y_i$.

   (e) Give a $(1-\alpha)$-confidence interval for $\mu$ based on $\hat{\mu}$ and on estimator of $\sigma^2$.

8. Consider the multivariate regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where the design matrix $\mathbf{X}$ is of size $n \times p$, and where the elements $e_i$ of the vector $\mathbf{e}$ are independent random variables with $\mathbb{E}\, e_i = 0$ and $\mathbb{V}\mathrm{ar}\, e_i = \sigma^2$. The least squares estimator of $\beta$ is given by $\hat{\beta} = \mathbf{LY}$, where $\mathbf{L} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Let $\tilde{\beta}$ be another linear estimator $\beta$, that is an estimator of the form $\tilde{\beta} = \mathbf{MY}$, where $\mathbf{M}$ is another non-random matrix of appropriate dimensions. We also assume that $\mathbf{MX} = \mathbf{I}_p$, where $\mathbf{I}_p$ is the $p$-dimensional unit matrix.

   (a) Show that $\tilde{\beta}$ is an unbiased estimator of $\beta$.

   (b) Show that $\mathbb{C}\mathrm{ov}(\tilde{\beta} - \hat{\beta}) = \sigma^2(\mathbf{MM}^\top - (\mathbf{X}^\top\mathbf{X})^{-1})$.

   (c) Suppose that one is interested in estimating the sum of the elements of $\beta$ only. Call this sum $\theta$ and notice that we can write $\theta = \mathbf{1}^\top\beta$, where $\mathbf{1}$ is a column vector whose elements are all equal to one. Show that $\mathbf{1}^\top\hat{\beta}$ and $\mathbf{1}^\top\tilde{\beta}$ are both unbiased estimators of $\theta$ with variances $\sigma^2\mathbf{1}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{1}$ and $\sigma^2\mathbf{1}^\top\mathbf{M}\mathbf{M}^\top\mathbf{1}$ respectively.

(d) Which of the two estimators in (c) is more accurate? (*Hint:* Compute $\mathbb{V}\mathrm{ar}\,(\tilde{\beta}-\hat{\beta}) = \mathbf{1}^\top\mathbb{C}\mathrm{ov}(\tilde{\beta}-\hat{\beta})\mathbf{1}$.)

9. Let $Y_1,\ldots,Y_n$ be independent random variables such that each $Y_i$ has a $N(\beta_0+\beta_1 x_i,\sigma^2)$ distribution. Here the $x_i$ are known real numbers (design variables) and $\beta_0$, $\beta_1$ and $\sigma^2$ are unknown parameters.

(a) Give the formula for the joint density function of $(Y_1,\ldots,Y_n)$.

(b) Assume for a while that $\sigma^2$ is known to us. Show that the maximum likelhood estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of $\beta_0$ and $\beta_1$ coincide with the least squares estimators in the context of linear regression.

(c) If we also consider $\sigma^2$ as unknown as well, show that the maximum likelihood estimator $\widehat{\sigma^2}$ of $\sigma^2$ is given by

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}(Y_k-\hat{Y}_k)^2,$$

where $\hat{Y}_k = \hat{\beta}_0+\hat{\beta}_1 x_i$.

(d) What is the distribution of $\sum_{k=1}^n(Y_k-\hat{Y}_k)^2$? Determine $c$ such that $c\,\widehat{\sigma^2}$ is an unbiased estimator of $\sigma^2$.

10. Let $X_1,\ldots,X_n$ be independent random variables with a common Gamma distribution and $n\geq 25$. This Gamma has density function (for $x>0$)

$$f_\theta(x) = C(\alpha)\theta^{-\alpha}x^{\alpha-1}e^{-x/\theta},$$

for some positive constants $\alpha,\theta$ and $C(\alpha)$.

(a) Show that $\frac{1}{C(\alpha)} = \int_0^\infty \theta^{-\alpha}x^{\alpha-1}e^{-x/\theta}\,dx$ and that $C(\alpha)$ doesn't depend on $\theta$.

(b) The constants $C(\alpha)$ have the property that $C(\alpha+1)=C(\alpha)/\alpha$. Show that $\mathbb{E}X_{24}=\alpha\theta$, $\mathbb{E}X_{24}^2=\alpha(\alpha+1)\theta^2$ and that $\mathrm{Var}\,X_{24}=\alpha\theta^2$.

(c) Suppose that $\alpha$ is known and that $\theta$ is an unknown parameter. Find the moment estimator of $\theta$.

(d) Show that the Fisher information $I(\theta)$ is equal to $\alpha/\theta^2$.

11. Let $X$ be a random variable for which $\mathbb{E}|X|^3 < \infty$. Then $\mathbb{E}X^3$, $\mathbb{E}X^2$ and $\mathbb{E}X$ are all well-defined and we can compute the *coefficient of skewness*

$$c(X) = \frac{\mathbb{E}(X-\mu)^3}{\sigma^3},$$

where $\mu = \mathbb{E}X$ and $\sigma = \sqrt{\operatorname{Var} X}$.

   (a) Show that $c$ is *scale and location invariant*, i.e. if $Y = aX + b$ ($a$ and $b$ are constants), then $c(Y) = c(X)$.

   (b) Show that $\mathbb{E}(X-\mu)^3 = \mathbb{E}X^3 - 3\mu\sigma^2 - \mu^3$.

   (c) Let $X$ have a density $f$, that is symmetric around $\mu$, so $f(\mu + x) = f(\mu - x)$ for all $x$. Show that $c(X) = 0$. (You may want to use $Y = X - \mu$).

   (d) Let $X$ be a $\chi^2_n$-distributed random variable. It is known that $\mathbb{E}X^3 = n(n+2)(n+4)$. Compute $c(X)$. How would you interpret the result for big values of $n$.

12. Let $X_1, \dots, X_n$ be a sample from a $N(\mu, \sigma^2)$ distribution. Assume that $\sigma^2$ is known. We consider the testing problem $H_0 : \mu = \mu_0$ against $H_A : \mu = \mu_A$ for some $\mu_A < \mu_0$ at significance level $\alpha$.

   (a) Show that the Neyman-Pearson test rejects the null hypothesis for small values of $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, say for $\overline{X} < x(\mu_0)$. Give an expression for $x(\mu_0)$.

   (b) What is the *uniformly most powerful test* for the testing problem $H_0 : \mu = \mu_0$ against $H_A : \mu < \mu_0$ at significance level $\alpha$?

   (c) Compute for the test of part (a) the *power* in $\mu = \mu_A$ for $\mu_A < \mu_0$, i.e. the probability that the test rejects the null hypothesis, when the mean of the normal distribution is equal to $\mu_A$. This probability depends on $n$. Compute the limit for $n \to \infty$.

   (d) Give an explicit expression for the confidence region (interval) $C = \{\mu : \overline{X} > x(\mu)\}$.

   (e) Suppose that $n = 100$, $\alpha = 0.10$, $\sigma^2 = 1$ and $\overline{X} = 3.14$ is found in a particular sample. Would the null hypothesis $H_0 : \mu = 0$ be rejected in favor of the alternative $\mu < 0$?

13. An urn with $N$ balls contains $r$ red ones (numbered from 1 to $r$) and the remaining $N - r$ are white. A random sample *without* replacement of size $n$ is drawn. The random variable $X_i$ ($i = 1, \dots, r$) is 1 if ball $i$ is in the sample and 0 otherwise. Let $X = X_1 + \cdots + X_r$.

(a) Show that for all $i$ we have $\mathbb{P}(X_i = 1) = \frac{n}{N}$, using binomial coefficients, and compute $\mathbb{E}\, X_i$ and $\mathbb{V}\mathrm{ar}\, X_i$.

(b) Compute $\mathbb{E}\, X$.

(c) Show that for all $i \neq j$ one has $\mathbb{P}(X_i = 1, X_j = 1) = \frac{n(n-1)}{N(N-1)}$, again using binomial coefficients.

(d) Let $i \neq j$. Show that $\mathbb{C}\mathrm{ov}(X_i, X_j) = -\frac{n(N-n)}{N^2(N-1)}$.

(e) Show that $\mathbb{V}\mathrm{ar}\, X = r\frac{n}{N}\frac{N-n}{N}\frac{N-r}{N-1}$.

(f) Suppose that $r, N \to \infty$ such that $\frac{r}{N} \to p$. Compute the limit of $\mathbb{V}\mathrm{ar}\, X$ and give an intuitive explanation for this result.

14. Let $X_1, \ldots, X_n$ be IID with common distribution function $F$ determined by $F(x) = 1 - \exp(-\lambda(x - \mu))$ for $x > \mu$, where $\lambda$ is known, but $\mu$ is an unknown parameter. Let $M = \min\{X_1, \ldots, X_n\}$ and $Y = n\lambda(M - \mu)$.

(a) Show that $\mathbb{P}(M > x) = \exp(-n\lambda(x - \mu))$ for $x > \mu$.

(b) Show that $Y$ has a standard exponential distribution (its density $f_Y$ is then $f_Y(y) = \exp(-y)$ for $y > 0$).

(c) Show that $\mathbb{E}\, M = \mu + \frac{1}{n\lambda}$ and that $\mathbb{V}\mathrm{ar}\, M = \frac{1}{n^2\lambda^2}$.

(d) We use $M$ as an estimator of $\mu$. Compute its mean squared error.

(e) As an alternative for $M$ we could use the moment estimator of $\mu$. What is this estimator and what is its mean squared error? Is this estimator a good alternative for $M$?

(f) Since we know that $\mu < M$, we compute a $1 - \alpha$ confidence interval for $\mu$ of the form $(M - c, M)$ for some $c > 0$. Show that $c = -\frac{\log \alpha}{n\lambda}$.

(g) Consider the one-sided hypothesis testing problem $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$ at significance level $\alpha$. Use $M$ as the test statistic. Show that $H_0$ is not rejected if and only if $\mu_0$ belongs to the confidence interval of part (f).

15. Let $X_1, \ldots, X_n$ be a random sample from a distribution which has density $f_\theta$ given by

$$f_\theta(x) = \frac{2}{\sqrt{\pi\theta}} e^{-x^2/\theta} 1_{(0,\infty)}(x),$$

where $\theta > 0$ is an unknown parameter.

(a) Show that $f_\theta$ is a density (on $(0, \infty)$).

(b) What is the density of $(X_1, \ldots, X_n)$?

(c) Show that the maximum likelihood estimator of $\theta$ is given by $\hat{\theta} = 2\sum_{i=1}^{n} X_i^2/n$.

(d) Let $Y$ have a normal $N(0, \sigma^2)$ distribution. Determine $\sigma^2$ such that $|Y|$ has density $f_\theta$.

(e) Show that $\mathbb{E}\, X_1^2 = \theta/2$. Is $\hat{\theta}$ an unbiased estimator of $\theta$?

(f) Show that the Fisher information $I(\theta)$ is equal to $1/2\theta^2$.

(g) Characterize the asymptotic distribution of $\hat{\theta}$ (for $n \to \infty$).

16. Let $Y$ be a random variable that has an exponential distribution with parameter $\lambda > 0$. Put $X = Y^2$. We observe the independent random variables $X_1, \ldots, X_n$, all of them having the same distribution as $X$.

(a) Compute $\mathbb{P}(X > x)$ for $x > 0$ and deduce that

$$f(x) = \frac{1}{2}\lambda x^{-1/2} e^{-\lambda x^{1/2}},\ x > 0$$

is a density of $X$.

(b) Show that the moment estimator of $\lambda$ is given by $\sqrt{\frac{2}{\bar{X}}}$.

(c) Use that $\mathbb{E}\, X = 2/\lambda^2$ and the law of large numbers to show consistency of the moment estimator.

(d) Show that the maximum likelihood estimator of $\lambda$ is given by $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i^{1/2}}$.

(e) Show that the Fisher information $I(\lambda) = 1/\lambda^2$.

(f) Characterize the asymptotic distribution of $\hat{\lambda}$.

(g) Show that $\left(\frac{\hat{\lambda}}{1+\frac{z(\alpha/2)}{\sqrt{n}}}, \frac{\hat{\lambda}}{1-\frac{z(\alpha/2)}{\sqrt{n}}}\right)$ is a $(1-\alpha)$-confidence interval for $\lambda$, where $z(\alpha/2)$ is the upper $\alpha/2$-quantile of the standard normal distribution: $\mathbb{P}(N(0,1) > z(\alpha/2)) = \alpha/2$.

17. We observe $X_1, \ldots, X_n$, independent random variables with a common Poisson distribution with parameter $\theta$.

(a) What is the probability mass function or frequency function of the random vector $(X_1, \ldots, X_n)$?

(b) Consider the hypotheses $H_0 : \theta = \theta_0$ and $H_A : \theta = \theta_1$, where $\theta_1 > \theta_0$. Show that the Neyman-Pearson test rejects the null hypothesis for "large values" of $S = \sum_{i=1}^{n} X_i$.

(c) Let $n = 100$, $\theta_0 = 1$ and $\theta_1 = 3$. Use the normal approximation to show that the critical value of the test is approximately 117 (we reject $H_0$ if $S \geq 117$), for significance level $\alpha = 0.05$.

(d) What is the power of this test in $\theta_1 = 3$? (Approximate again with a normal distribution.)

(e) Would the test of the previous part change if we take instead $\theta_1 = 5$ and keeping $\theta_0 = 1$, $\alpha = 0.05$?

(f) Is the test uniformly most powerful for testing $H_0$ against the alternative $\theta > 1$?

18. Consider the multivariate linear regression model $Y = X\beta + \varepsilon$, where $Y$ is a random $n$-vector, $X$ a $(n \times p)$-matrix (non-random), $\beta$ a $p$-dimensional parameter vector, $\varepsilon$ a random $n$-vector with $\mathbb{E}\,\varepsilon = 0$ and $\mathbb{C}\text{ov}(\varepsilon) = \sigma^2 I_n$ ($I_n$ is the $n \times n$ identity matrix. We study the least squares estimator $\hat{\beta}$ of $\beta$. By definition it is the minimizer over $\beta$ of $SS(\beta) = \varepsilon^\top \varepsilon$. We need the matrices $L = (X^\top X)^{-1} X^\top$ and $P = XL$ (we assume that $X^\top X$ is invertible).

(a) Show that $X^\top (I - P) = 0$. Write $Y - X\beta = (I - P)Y + X(LY - \beta)$ to show that

$$SS(\beta) = Y^\top (I - P)^\top (I - P)Y + (\beta - LY)^\top X^\top X(\beta - LY).$$

(b) Deduce that $\hat{\beta} = LY$. What is $\mathbb{E}\,Y$? Is $\hat{\beta}$ an unbiased estimator of $\beta$?

(c) Let $e = Y - X\hat{\beta}$. Show that $e = (I - P)\varepsilon$.

(d) Show that $(I - P)^\top (I - P) = I - P$ and conclude that $\mathbb{C}\text{ov}(e) = \sigma^2 (I_n - P)$

(e) Show that $SS(\hat{\beta}) = e^\top e$. Show then that the expected value of $SS(\hat{\beta})$ is equal to $\sigma^2(n - p)$. (*Hint:* write $SS(\hat{\beta}) = \text{tr}\,(ee^\top)$ and use also somewhere else the rule $\text{tr}(AB) = \text{tr}(BA)$.)

19. Let $X = (X_1, X_2)^\top$ be a vector of independent random variables that both have a normal $N(0, \sigma^2)$ distribution ($\sigma^2 > 0$). Let $Y = (Y_1, Y_2)^\top$ with $Y = AX$, where $A$ is the matrix

$$A = \begin{pmatrix} a & -1 \\ b & ab \end{pmatrix},$$

for real numbers $a$ and $b$ ($b \neq 0$).

(a) Compute the covariance matrix of $Y$.

(b) What is the distribution of $Y$?

(c) Show that $Y_1$ and $Y_2$ are independent random variables.

(d) Show that $Y_1^2$ and $Y_2^2$ are independent random variables.

(e) For certain real constants $\lambda_1$ and $\lambda_2$ put $U = \lambda_1 Y_1^2 + \lambda_2 Y_2^2$. How do we have to choose $\lambda_1$ and $\lambda_2$ such that $U$ has a $\chi_2^2$-distribution?

(f) How to choose the constants $\lambda_1$ and $\lambda_2$ in the previous part such that $U$ has an exponential distribution with parameter 1?

20. We observe $X_1, \ldots, X_n$, independent random variables with a common exponential distribution depending on a parameter $\theta > 0$ with density

$$f(x|\theta) = \frac{1}{\theta} e^{-x/\theta},\ x > 0.$$

(a) What is the probability density function of the random vector $(X_1, \ldots, X_n)$?

(b) Suppose that we already know that $U = \sum_{k=1}^{n-1} X_k$ has a Gamma distribution with density

$$f_{n-1}(u|\theta) = \frac{u^{n-2}}{\theta^{n-1}(n-2)!} e^{-u/\theta},\ u > 0.$$

Show by computing the convolution integral that $S = \sum_{k=1}^{n} X_k = U + X_n$ has density

$$f_n(s|\theta) = \frac{s^{n-1}}{\theta^n (n-1)!} e^{-s/\theta},\ s > 0.$$

N.B.: Also $S$ thus has a gamma distribution.

(c) Show that $S/\theta$ has density

$$f_n(s|1) = \frac{s^{n-1}}{(n-1)!} e^{-s}.$$

(d) Consider the hypotheses $H_0 : \theta = \theta_0$ and $H_A : \theta = \theta_1$, where $\theta_1 > \theta_0$. Show that the Neyman-Pearson test rejects the null hypothesis for "large values" of $S$, $S > c$ say.

(e) If $\alpha$ is the significance level of the test, show that $c = \theta_0 \gamma_\alpha$, where $\gamma_\alpha$ satisfies $\int_{\gamma_\alpha}^{\infty} f_n(s|1)\, ds = \alpha$.

(f) The power of this test in $\theta_A$ is $\pi(\theta_A) = \mathbb{P}(S > \theta_0 \gamma_\alpha | \theta_A)$. Compute $\lim_{n\to\infty} \pi(\theta_A)$.

(g) Is the Neyman-Pearson test uniformly most powerful for testing $H_0$ against the alternative $\theta > \theta_0$?

21. Consider the multivariate regression model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$, where the design matrix $X$ is of size $n \times p$, and where the elements $e_i$ of the vector $\mathbf{e}$ are independent random variables with $\mathbb{E}\, e_i = 0$ and $\mathbb{V}\mathrm{ar}\, e_i = \sigma^2$. The least squares estimator of $\beta$ is given by $\hat{\beta}_n = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$.

   (a) Suppose that one has an additional (row) vector of design variables $x_{n+1}$. The corresponding response variable $Y_{n+1}$ is then predicted by $\hat{Y}_{n+1|n} = x_{n+1}\hat{\beta}_n$. Let $\varepsilon_{n+1|n} = \hat{Y}_{n+1|n} - Y_{n+1}$ be the prediction error. Why are $\hat{Y}_{n+1|n}$ and $Y_{n+1}$ independent?

   (b) Compute the expectation $\mathbb{E}\, \varepsilon_{n+1|n}$ and the variance $\mathbb{V}\mathrm{ar}\, \varepsilon_{n+1|n}$.

   (c) If we also observe $Y_{n+1}$ we can compute a new least squares estimator $\hat{\beta}_{n+1}$ following the usual least squares procedure, but now based on $n+1$ observations. It turns out that the following recursive relationship holds

$$\hat{\beta}_{n+1} = \hat{\beta}_n + \frac{1}{1+d} (\mathbf{X}^\top \mathbf{X})^{-1} x_{n+1}^\top (Y_{n+1} - x_{n+1}\hat{\beta}_n),$$

   where $d = x_{n+1} (\mathbf{X}^\top \mathbf{X})^{-1} x_{n+1}^\top$. Using the estimator $\hat{\beta}_{n+1}$, we predict $Y_{n+1}$ by $\hat{Y}_{n+1} = x_{n+1}\hat{\beta}_{n+1}$. Show that

$$\hat{Y}_{n+1} = \frac{1}{1+d} \hat{Y}_{n+1|n} + \frac{d}{1+d} Y_{n+1}.$$

   (d) Let $\varepsilon_{n+1}$ be the associated prediction error, $\varepsilon_{n+1} = \hat{Y}_{n+1} - Y_{n+1}$. Compute $\mathbb{E}\, \varepsilon_{n+1}$ and $\mathbb{V}\mathrm{ar}\, \varepsilon_{n+1}$.

   (e) Which of the two predictors $\hat{Y}_{n+1|n}$ and $\hat{Y}_{n+1}$ would you prefer?

   (f) Suppose that we also know that the $e_i$ are $N(0, \sigma^2)$ distributed random variables with *unknown* $\sigma^2$. Show that $x_{n+1}\hat{\beta}_{n+1}$ has a $N(x_{n+1}\beta, \frac{d\sigma^2}{1+d})$ distribution.

   (g) Let $R = \sum_{i=1}^{n+1} (Y_i - x_i \hat{\beta}_{n+1})^2$. It is known that $\frac{R}{\sigma^2}$ has a $\chi^2_{n+1-p}$ distribution. Show that

$$T := \frac{x_{n+1}(\hat{\beta}_{n+1} - \beta)}{\sqrt{\frac{R}{n+1-p} \frac{d}{1+d}}}$$

   has a $t_{n+1-p}$-distribution.

   (h) Construct a $(1-\alpha)$-confidence interval for $x_{n+1}\beta$ based on $\hat{\beta}_{n+1}$.

22. (a) Let $X_1$ and $X_2$ be independent random variables, both with a geometric distribution. Let $Y = X_1 + X_2$. Give an expression for $\mathbb{P}(Y = k)$ and deduce that $Y$ has a negative binomial distribution with $r = 2$.

(b) Let $X_1, \ldots, X_r$ be independent random variables, all with a common geometric distribution with parameter $p$. It can be shown that $Y = X_1 + \cdots + X_r$ has a negative binomial distribution with parameters $r$ and $p$. It can also be shown that $\mathbb{V}\mathrm{ar}\, X_1 = (1-p)/p^2$. Compute expectation and variance of $Y$.

(c) Let $f(k|p)$ be the probability mass function, $f(k|p) = \mathbb{P}(Y = k|p)$, and $\dot{l}(p) = \frac{\partial}{\partial p} \log f(Y|p)$. Show that $\mathbb{E}\, \dot{l}(p) = 0$. Let $I(p) = \mathbb{E}\, \dot{l}(p)^2$. Compute $I(p)$.

Let $Y_1, \ldots, Y_n$ be a sample from a negative binomial distribution with parameters $p$ (unknown) and $r$.

(d) Compute the maximum likelihood estimator $\hat{p}_n$ of $p$ and show that it is equal to the moment estimator.

(e) What is the asymptotic distribution of $\sqrt{n}(\hat{p}_n - p)$?

(f) Suppose that $n = 100$, $r = 4$, and that the sample is such that $\hat{p}_{100} = 0.75$. Give an approximate 95% confidence interval for $p$.

23. Let $X_1, \ldots, X_n$ be sample from a Poisson distribution with parameter $\lambda$ (unknown). Let $T = X_1 + \cdots + X_n$.

(a) Consider the simple hypothesis testing problem $H_0 : \lambda = \lambda_0$ against $H_A : \lambda = \lambda_1$, where $\lambda_0 > \lambda_1$. Show that the Neyman-Pearson test rejects $H_0$ for 'small values' of $T$, $T \leq c_n$ say, for some integer $c_n$.

(b) Show that the function $\lambda \mapsto \mathbb{P}(T \leq c_n|\lambda)$ is decreasing. *Hint:* Let $\lambda_1 < \lambda_2$, and let $U$ have a Poisson distribution with parameter $\lambda_1$ and $V$, independent of $U$, have a Poisson distribution with parameter $\lambda_2 - \lambda_1$. Use the trivial inequality $U + V \geq U$.

(c) Let $\alpha = \mathbb{P}(T \leq c_n|\lambda_0)$. Consider the composite testing problem $H_0 : \lambda \geq \lambda_0$ against $H_A : \lambda < \lambda_0$. We use (again) the test that rejects $H_0$ if $T \leq c_n$. Compute $\sup_{\lambda \geq \lambda_0} \mathbb{P}(T_n \leq c_n|\lambda)$, and deduce that this test has significance level $\alpha$.

(d) Is the above test uniformly most powerful for the testing problem under consideration?

(e) Replace $c_n$ with $\xi_n := n\lambda_0 + \xi\sqrt{n\lambda_0}$ for some (negative) real number $\xi$. Compute, by using the Central Limit Theorem (CLT) and in terms of the cdf $\Phi$, $\mathbb{P}(T \leq \xi_n | \lambda_0)$. How should one choose $\xi$ to have the last probability (approximately) equal to $\alpha$?

(f) Suppose that $n = 100$, $\lambda_0 = 1$. Give a numerical value for $\xi_n$, if $\alpha = 0.0202$. If $T = 90$ is observed, should one reject $H_0$?

(g) Fix some $\lambda_1 < \lambda_0$. Use the CLT again to show that the (asymptotic) power of the test, $\mathbb{P}(T \leq \xi_n | \lambda_1)$ is equal to $\Phi(\xi\sqrt{\lambda_0/\lambda_1} + (\lambda_0 - \lambda_1)\sqrt{n/\lambda_1})$. What happens with this probability as $n \to \infty$?

24. In this exercise we consider *quadratic* regression, we assume a model of the form $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$, $i = 1, \ldots, n$. The $e_i$ are assumed to be *iid* with a common normal $N(0, \sigma^2)$ distribution. In matrix notation, we summarize the model by writing

$$\mathbf{Y} = X\beta + \mathbf{e},$$

following the usual conventions.

(a) How would you cast this model as an ordinary *linear* regression model by choosing the right independent variables?

(b) We know that it is important that $X$ has rank 3. Show that this is the case if at least three of the $x_i$ ($x_1, x_2, x_3$ for instance) are different. *Hint:* compute the determinant

$$\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix}.$$

(c) Show that the rank of $X$ is at most 2, if all the $x_i$ assume at most two different values.

(d) Let $\hat{\beta}$ be the least squares estimator of $\beta$ and $\hat{\mathbf{Y}} = X\hat{\beta}$. Show that $\mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Q}\mathbf{e}$, where $\mathbf{Q} = I - X(X^\top X)^{-1}X^\top$. Show also that $\mathbf{Q}^2 = \mathbf{Q}$.

(e) Determine a matrix $\mathbf{M}$ such that

$$\mathbf{U} := \begin{pmatrix} \mathbf{Y} - \hat{\mathbf{Y}} \\ \hat{\beta} - \beta \end{pmatrix} = \mathbf{M}\mathbf{e}.$$

(f) Show that $\mathbf{U}$ has covariance matrix equal to $\sigma^2 \begin{pmatrix} \mathbf{Q} & 0 \\ 0 & (X^\top X)^{-1} \end{pmatrix}$. Are $\mathbf{Y} - \hat{\mathbf{Y}}$ and $\hat{\beta}$ independent?

(g) Let $S^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}})$. It is known that $\frac{S^2}{\sigma^2}$ has a $\chi^2$-distribution with $n - 3$ degrees of freedom. Use this to deduce that

$$\frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}}$$

has a $t$-distribution with $n-3$ degrees of freedom, where $s_{\hat{\beta}_i} = S\sqrt{(X^\top X)^{-1}_{ii}}$.

(h) Suppose that $n = 20$ and that computations with the data result in $\hat{\beta}_2 = s_{\hat{\beta}_2} = 0.42$. Give a 95% confidence interval for $\beta_2$.

(i) Suppose that one wants to test the hypothesis that the regression is linear in one variable. Formulate this as a testing problem on the coefficients $\beta_i$. Should one reject this hypothesis in the situation of the previous part?

25. Consider a two-dimensional random vector $(X, Y)$ which has a density $f$ on the square $(0, 1) \times (0, 1)$ given by

$$f(x, y) = \begin{cases} \frac{4y}{x} \text{ if } y < x \\ 0 \text{ else }. \end{cases}$$

(a) Show that the marginal densities of $X$ and $Y$ are given by $f_X(x) = 2x$ and $f_Y(y) = -4y \log y$, for $x, y \in (0, 1)$.

(b) Compute $\mathbb{E}\, X$ and $\mathbb{V}\mathrm{ar}\, X$.

(c) Show that $y \mapsto y^k \log y$ has $y \mapsto \frac{1}{k+1} y^{k+1} \log y - \frac{1}{(k+1)^2} y^{k+1}$ as a primitive function ($k \geq 0$).

(d) Compute $\mathbb{E}\, Y$ and $\mathbb{V}\mathrm{ar}\, Y$.

(e) Compute $\mathbb{E}\, XY$, $\mathbb{C}\mathrm{ov}(X, Y)$ and the correlation coefficient.

26. Consider the simple regression model $Y_i = \beta_0 + \beta_1 x_i + e_i$ ($i = 1, \ldots, n$). Assume that the $e_i$ are independent with a common $N(0, \sigma^2)$ distribution. Let $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ be the model in matrix form. Assume that in a certain experiment $n = 42$ and that the matrix

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} 0.03 & -0.015 \\ -0.015 & 0.04 \end{pmatrix}.$$

Other relevant statistics are $||\mathbf{Y} - \hat{\mathbf{Y}}||^2 = 160$ and the least squares estimators become $\hat{\beta}_0 = 1.90$ and $\hat{\beta}_1 = 0.65$.

(a) Compute 95% confidence intervals for $\beta_0$ and $\beta_1$.

(b) Consider the testing problem $H_0 : \beta_0 = 2$ versus $H_A : \beta_0 \neq 2$. Will the null hypothesis be rejected at a significance level of $\alpha = 5\%$? And if $\alpha = 1\%$?

(c) Under the previous null hypothesis it holds that $\mathbb{P}(\hat{\beta}_0 > 1.90) = 0.61$ and that $\mathbb{P}(\hat{\beta}_0 < 1.90) = 0.39$. For which values of $\alpha$ would one reject this null hypothesis with the given data?

27. Consider an *iid* sequence $X_1, X_2, \ldots$ with a common density $f_\theta(x) = \theta x \exp(-\frac{1}{2}\theta x^2)$, for $x > 0$ and $\theta > 0$.

(a) Let $\mu_k = \mathbb{E}\, X^k$. Use integration by parts to show that $\mu_k = \frac{k}{\theta}\mu_{k-2}$ for $k \geq 2$. Compute $\mu_2$, $\mu_4$ and $\mathbb{V}\mathrm{ar}\,(X_1^2)$.

(b) What is the joint density of $(X_1, \ldots, X_n)$?

(c) Compute the maximum likelihood estimator $\hat{\theta}_n$ of $\theta$ based on the observations $X_1, \ldots, X_n$.

(d) What is the Fisher information $I(\theta)$ in one observation?

(e) What is the asymptotic distribution of $\sqrt{n}(\frac{\hat{\theta_n}}{\theta} - 1)$?

(f) Show that $(\frac{\hat{\theta}_n}{1+z_{\alpha/2}/\sqrt{n}}, \frac{\hat{\theta}_n}{1-z_{\alpha/2}/\sqrt{n}})$ is an approximate $1-\alpha$ confidence interval for $\theta$. ($z_{\alpha/2}$ is such that $\Phi(z_{\alpha/2}) = 1 - \alpha/2$)

(g) Consider the following testing problem: $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_A$ at significance level $\alpha$, with $\theta_A > \theta_0$. Show that the most powerful test rejects the nulhypothesis for small values of $\sum_{k=1}^n X_k^2$.

(h) One can show that $\frac{\theta_0}{2}\sum_{k=1}^n X_k^2$ has a Gamma distribution with parameters $n$ and 1 under the null hypothesis. Use this to describe the critical region for the most powerful test of the previous part.

(i) Is this most powerful test uniformly most powerful for the testing problem $H_0 : \theta = \theta_0$ versus $H_A : \theta > \theta_0$? Is it also uniformly most powerful for the testing problem $H_0 : \theta = \theta_0$ versus $H_A : \theta < \theta_0$?

28. Consider two independent random variables $U$ and $Y$. Assume that $U$ has a Bernoulli distribution with parameter $p$ and that $Y$ has an exponential distribution with parameter $\lambda$. Let $X = UY$ and let $F$ be the distribution function of $X$.

(a) Compute $F(x)$ for all $x \in \mathbb{R}$.

(b) Show that $\mathbb{E}\, X = \frac{p}{\lambda}$ and $\mathbb{E}\, X^2 = \frac{2p}{\lambda^2}$.

Consider a sample $X_1, \ldots, X_n$, all having distribution function $F$. The parameters $p$ and $\lambda$ are not known and are to be estimated. Let $\hat{p}_n$ and $\hat{\lambda}_n$ be their moment estimators.

(c) Write down the two equations for the moment estimators of $p$ and $\lambda$.

(d) Show that

$$\begin{aligned}\hat{p}_n &= \frac{2(\overline{X}_n)^2}{\overline{X^2}_n}\\ \hat{\lambda}_n &= \frac{2\overline{X}_n}{\overline{X^2}_n}.\end{aligned}$$

(e) What are the limits in probability of $\overline{X}_n$ and $\overline{X^2}_n$ as $n \to \infty$?

(f) Are the two moment estimators consistent?

(g) Let $N$ be the number of zero observations. Show that $N$ has a Binomial distribution with parameters $n$ and $1-p$.

(h) Give an approximate $100(1-\alpha)\%$-confidence interval for $p$ based on $N$.

29. Consider a regression model $Y_i = x_i\beta + e_i$, where the design (row) vector $x_i$ has a very special form. This form is motivated by the following (thought) experiment. A subject $i$ is classified according to some criterion in a situation where $p$ disjoint categories are available. One sets $x_{ij} = 1$ if $i$ falls into category $j$ and zero otherwise. The row vector $x_i$ that represents the classification has exactly one 1 (it is on place $j$ iff $i$ falls into the $j$-th category) and all the remaining elements are 0. In other words, it is a unit vector. The theoretical response of an individual $i$ depends on the category to which it belongs and is $\beta_j$ if $x_{ij} = 1$. Responses are measured with normally distributed errors $e_i$ having expectation zero and variance $\sigma^2$. Assume that there $n$ subjects considered, so that the *vector* $\mathbf{Y}$ of responses is $n$-dimensional. The design matrix $\mathbf{X}$ is built by stacking the row vectors $x_i$ one underneath the other. The usual assumptions on the regression are assumed to be in force.

(a) Let $n_j$ be the number of subjects that fall into category $j$. Show that $\mathbf{X}^\top\mathbf{X}$ is a diagonal matrix with $jj$-element equal to $n_j$.

(b) Let $\hat{\beta}$ be the ordinary least squares estimator. Show that the elements of $\hat{\beta}$ are independent random variables.

(c) In this case there is a simple explicit expression for each of the $\hat{\beta}_j$. Give it and interpret the result.

(d) Suppose that one is interested in the testing of the null hypothesis $H_0 : \beta_1 + \beta_2 = 0$ against the alternative $H_A : \beta_1 + \beta_2 \neq 0$. Let $s^2$ be the usual estimator of $\sigma^2$. Show that

$$T := \frac{\hat{\beta}_1 + \hat{\beta}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has a $t$-distribution under the null hypothesis. What is the numbers of degrees of freedom?

(e) Construct a $100(1-\alpha)\%$-confidence interval for $\beta_1 + \beta_2$ based on $T$.

(f) Suppose that for a particular set of observations the numerical values of the two limits of the confidence interval turn out to be $-0.13$ and $+0.87$ with $\alpha = 0.10$. Would you reject the null hypothesis at the significance level $\alpha$?

(g) Using the same confidence interval, would you reject the null hypothesis at a smaller significance level?

30. In a sequence of Bernoulli experiments one writes down the number of trials needed to obtain the first success, call it $X$, and the number of additional trials needed to obtain the second success, call it $Y$.

(a) Show that the joint distribution of $(X, Y)$ is given by

$$\mathbb{P}(X = k, Y = m) = p^2(1-p)^{k+m-2}, \text{ for } k, m = 1, 2, \ldots$$

Are $X$ and $Y$ independent?

(b) Show that the Fisher information is given by $I(p) = \frac{2}{p^2(1-p)}$. (You may use that $\mathbb{V}\mathrm{ar}\, X = \frac{1-p}{p^2}$.)

(c) Suppose one independently repeats $n$ times the above procedure, resulting in observations $X_1, Y_1, \ldots, X_n, Y_n$. Show that the maximum likelihood $\hat{p}$ based on these observations is equal to $\frac{1}{\frac{1}{2}(\bar{X}+\bar{Y})}$.

(d) Characterize the asymptotic distribution of $\hat{p}$.

31. Let $X_1, \ldots, X_n$ be *iid* random variables with a common distribution having expectation $\mu$, variance $\sigma^2$ and continuous distribution function $F$. Let $A_1, \ldots, A_n$ be *iid* random variables having a Bernoulli distribution with parameter $p$. Assume that the $A_i$ are independent from the $X_i$. Put $Y_i =$

$A_iX_i$ and denote by $G$ the (common) distribution function of the $Y_i$. We assume that we observe only the $Y_i$. Let $U_i = I_{\{Y_i \neq 0\}}$, the indicator of the event $\{Y_i \neq 0\}$ and $S_n = \sum_{i=1}^n U_i$.

(a) What is $\mathbb{P}(Y_i = 0)$? Express $G(y)$ in terms of $F$ and $p$. Distinguish in your calculations between $y < 0$, $y = 0$ and $y > 0$.

(b) Show that $\mathbb{E}\, Y_i = p\mu$ and $\mathbb{E}\, Y_i^2 = p(\sigma^2 + \mu^2)$. What is $\mathbb{V}\mathrm{ar}\, Y_i$?

(c) What is $\mathbb{E}\,(Y_i - \mu U_i)$? Compute $\mathbb{V}\mathrm{ar}\,(Y_i - \mu U_i)$.

(d) Suppose that $p$ is known. Find an estimator of $\mu$ based on the method of moments. Is it an unbiased estimator?

(e) Characterize the limit distribution of this estimator for $n \to \infty$.

(f) In the rest of this exercise we suppose that $p$ is unknown. Give a consistent estimator of $p$ (you may omit an explanation of your answer).

(g) The estimator of part (d) is now useless for estimating $\mu$ and instead we will estimate $\mu$ by $\hat{\mu} := \sum_{i=1}^n Y_i/S_n$, provided $S_n > 0$, and by zero otherwise. Show that $\mathbb{P}(S_n = 0) \to 0$ for $n \to \infty$.

(h) Show that, assuming $S_n > 0$,

$$\sqrt{n}(\hat{\mu} - \mu) = \frac{n}{S_n} \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mu U_i).$$

(i) What is the limit distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mu U_i)$ for $n \to \infty$. Show that the limit distribution of $\sqrt{n}(\hat{\mu}_n - \mu)$ is normal with mean zero and variance $\sigma^2/p$.

(j) Suppose that one also wants to use $\hat{\mu}$ as an estimator of $\mu$, when $p$ is known. Is this to be preferred over using the moment estimator of part (d)?

32. Consider a sequence of *iid* random variables $X_1, X_2, \ldots$ whose distribution is given by $\mathbb{P}(X_1 = 0) = p$, $\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 1) = \frac{1}{2}(1-p)$. Let $\hat{p}_n$ be the maximum likelihood estimator of $p$ based on $n$ observations.

    (a) Show that $\mathbb{P}(X_1 = x) = p^{1-|x|}(\frac{1}{2}(1-p))^{|x|}$ for $x \in \{-1, 0, 1\}$. Give a formula for $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$.

    (b) Compute $\hat{p}_n$.

    (c) To compute a moment estimator, you have the choice between working with $\mathbb{E}\, X$ and with $\mathbb{E}\, X^2$. Make your choice and compute the corresponding estimator.

(d) Let $\ell(p)$ be the log-likelihood based on the single observation $X_1$. Show that $\frac{\partial}{\partial p}\ell(p) = \frac{1}{p} - \frac{|X_1|}{p(1-p)}$ and that this has expectation zero.

(e) Compute the Fisher information $I(p)$ in one observation.

(f) Characterize the asymptotic distribution of $\hat{p}_n$ for $n \to \infty$.

(g) Give a 95% (approximate) confidence interval for $p$. Give the numerical value when $n = 100$ with $\hat{p}_{100} = 0.2$.

(h) Suppose you have to test the hypothesis $H_0 : p = 0.25$ against the alternative $H_A : p \neq 0.25$ at the 5% significance level. If you take $\hat{p}_n$ as a test statistic, do you reject $H_0$?

33. Consider the following two regression models in vector form.

$$\mathbf{Y}_1 = \mathbf{X}\beta + \mathbf{e}_1$$
$$\mathbf{Y}_2 = \mathbf{X}\beta + \mathbf{e}_2.$$

In both models the matrix $\mathbf{X}$ of independent variables is the same, it is of size $n \times p$ and has rank $p$. Also the parameter vector $\beta \in \mathbb{R}^p$ is the same for both models. The random noise vector $\mathbf{e}_1$ has independent normally distributed elements with mean zero and variance $\sigma_1^2$, whereas the random noise vector $\mathbf{e}_2$ has independent normally distributed elements with mean zero and variance $\sigma_2^2$. The vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are independent too. For the two models we estimate $\beta$ separately by the least squares method, resulting in two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. The covariance matrices of these estimators are given by the usual formulas. Let $a$ be some real number and consider the 'mixed' estimator of $\beta$ given by the convex combination $\hat{\beta}(a) = a\hat{\beta}_1 + (1-a)\hat{\beta}_2$.

(a) Show that $\hat{\beta}(a)$ is an unbiased estimator of $\beta$.

(b) Let $f(a) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$. Let $a_0$ be the value of $a$ where $f$ is minimal. Compute $a_0$ (it will depend on $\sigma_1^2$ and $\sigma_2^2$) and show that $f(a_0) = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$.

(c) A second order Taylor expansion of $f$ is *exact* since $f$ is quadratic and it holds that $f(a) = f(a_0) + (\sigma_1^2+\sigma_2^2)(a-a_0)^2$. Let $\Sigma_a$ be the covariance matrix of $\hat{\beta}(a)$. Show that

$$\Sigma_a = \Sigma_{a_0} + (\sigma_1^2 + \sigma_2^2)(a - a_0)^2(\mathbf{X}^\top\mathbf{X})^{-1}.$$

(d) Suppose that $\lambda \in \mathbb{R}^p$ is a known vector and that one wants to estimate the parameter $\theta$ defined as $\lambda^\top\beta$. Show that $\lambda^\top\hat{\beta}(a)$ is an unbiased estimator of $\theta$. What is it's variance? Which of the estimators $\lambda^\top \hat\beta(a)$ (for $a \in \mathbb{R}$) would you prefer to estimate $\theta$? Which among the $\hat\beta(a)$ would you prefer to estimate $\beta$?

(e) The problem with this best 'estimator' is that it involves $a_0$ which depends on unknown parameters. Propose an estimator of $a_0$.

(f) One wants to test the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_A : \sigma_1^2 \neq \sigma_2^2$. Let $S_1^2$ be the usual unbiased estimator of $\sigma_1^2$ based on the observations $\mathbf{Y}_1$ and $S_2^2$ the companion estimator of $\sigma_2^2$ based on the observations $\mathbf{Y}_2$. Under the null hypothesis the ratio $R := S_1^2/S_2^2$ has a so-called $F_{n-p,n-p}$ distribution, see Table 5, as well as $S_2^2/S_1^2$. The test rejects $H_0$ with significance level $\alpha$ if $R < c_1$ or $R > c_2$ with $\mathbb{P}_{H_0}(R < c_1) = \mathbb{P}_{H_0}(R > c_2) = \alpha/2$. Show that $c_1 = \frac{1}{c_2}$.

(g) Assume that $n = 50$, $p = 10$, $\alpha = 0.10$ and that from the two samples one computes the value $R = 1.61$. Does the test reject $H_0$?

34. Let $f$ be the density of some random variable $Z$ and assume that $p := \mathbb{P}(Z > 0) \in (0, 1)$. Let $g$ be the joint density of a random vector $(X, Y)$ defined by

$$g(x,y) = \begin{cases} \frac{1}{p} f(x)f(y) & \text{if } x, y > 0 \\ \frac{1}{1-p} f(x)f(y) & \text{if } x, y < 0 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Show that the marginal density $g_X$ of $X$ is equal to $f$. (In your computation of $g_X(x)$ you distinguish between $x > 0$ and $x < 0$. The case $x = 0$ can be ignored.) What is the marginal density of $Y$?

(b) In what follows, we let $Z^+ = \max\{Z, 0\}$ and $Z^- = \max\{-Z, 0\}$. Note that $Z = Z^+ - Z^-$. (For example, if $Z = -3$, we find $Z^+ = 0$ and $Z^- = 3$.) Show that $\mathbb{E}\, X = \mathbb{E}\, Z$ and that

$$\mathbb{E}\,(XY) = \frac{(\mathbb{E}\, Z^+)^2}{p} + \frac{(\mathbb{E}\, Z^-)^2}{1-p}.$$

(c) Show that $\mathbb{C}\mathrm{ov}(X, Y) = (\sqrt{\frac{1-p}{p}}\, \mathbb{E}\, Z^+ + \sqrt{\frac{p}{1-p}}\, \mathbb{E}\, Z^-)^2$.

(d) According to the previous item $\mathbb{C}\mathrm{ov}(X, Y) > 0$. Argue why this is intuitively obvious.

(e) Suppose that $Z$ is standard normal. What are the marginal distributions of $X$ and $Y$? Are $X$ and $Y$ independent? Is $(X, Y)$ bivariate normal?

35. Let $X_1, \ldots, X_n$ be an *iid* sequence of Bernoulli random variables with probability $p$ on 'success'. If $p \in [0,1]$ is unknown, the Maximum Likelihood Estimator (MLE) of $p$ is the average $\overline{X}_n$. In this exercise it is *known* that $p > \frac{1}{2}$, but unknown otherwise and we try again to estimate $p$ by Maximum Likelihood, taking the information $p > \frac{1}{2}$ into account. As in the usual situation the log-likelihood is

$$\ell(p) = n\big(\overline{X}_n \log p + (1 - \overline{X}_n)\log(1-p)\big).$$

To obtain the MLE, we will maximize $\ell(p)$ over $p$ in the *closed* interval $[\frac{1}{2}, 1]$. We investigate the properties of the resulting MLE in this unusual situation. We will see that it is important to distinguish between the cases $\overline{X}_n \leq \frac{1}{2}$ and $\overline{X}_n > \frac{1}{2}$.

(a) Compute $\dot{\ell}(p) = \frac{\partial}{\partial p}\ell(p)$ and show that $\ell(p)$ is a decreasing function of $p$ on the interval $[\frac{1}{2}, 1]$ if $\overline{X} \leq \frac{1}{2}$. For which value of $p$ is $\ell(p)$ maximal in this case?

(b) If $\overline{X}_n > \frac{1}{2}$, compute the value of $p$ where $\ell(p)$ is maximal.

(c) Let $\hat{p}_n$ be the MLE of $p$. Conclude that $\hat{p}_n = \max\{\overline{X}_n, \frac{1}{2}\}$.

(d) Show that $\hat{p}_n$ is a consistent estimator of $p$.

(e) Show that $\mathbb{P}(\overline{X}_n \leq \frac{1}{2}) \leq \mathbb{P}(|\overline{X} - p| \geq p - \frac{1}{2})$ for $p > \frac{1}{2}$. Use Chebychev's inequality to show that $\mathbb{P}(\overline{X}_n \leq \frac{1}{2}) \to 0$ as $n \to \infty$ for $p > \frac{1}{2}$.

The Fisher information $I(p)$ is as usual equal to $\frac{1}{p(1-p)}$. We want to investigate whether the property

$$\mathbb{P}(\sqrt{nI(p)}(\hat{p}_n - p) \leq x) \to \Phi(x) \tag{1}$$

continues to hold for our estimation problem.

(f) Show that one has

$$\begin{aligned}\mathbb{P}(\sqrt{nI(p)}(\hat{p}_n - p) \leq x) = \; & \mathbb{P}(\sqrt{nI(p)}(\frac{1}{2} - p) \leq x, \overline{X}_n \leq \frac{1}{2}) \\ & + \mathbb{P}(\sqrt{nI(p)}(\overline{X}_n - p) \leq x) \\ & - \mathbb{P}(\sqrt{nI(p)}(\overline{X}_n - p) \leq x, \overline{X}_n \leq \frac{1}{2}).\end{aligned}$$

(g) Compute, taking into account that $p > \frac{1}{2}$ and use the Central limit theorem where needed, the limits of each of the three probabilities on the right in the above display. Does the convergence in Equation (1) hold true?

36. Consider the model for multiple regression, $Y = \tilde{X}\tilde{\beta} + e$. Assume that $Y$ and $e$ are $n$-dimensional random vectors, $\tilde{\beta} \in \mathbb{R}^p$ and $\tilde{X} \in \mathbb{R}^{n\times p}$. Accept the odd notation with the tildes for a while. As usual, we assume that the elements of $e$ are independent random variables with a common $N(0, \sigma^2)$ distribution. Suppose that $p = 3$, $n \geq 3$ and that

$$\tilde{X} = \begin{pmatrix} 1 & x_1 & 2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & 2 \end{pmatrix},$$

with $x_1 \neq x_2$ and $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2)^\top$.

(a) What is the rank of $\tilde{X}$. Can we compute the ordinary least squares estimator of $\tilde{\beta}$?

(b) Show that we can write $\tilde{X}\tilde{\beta} = X\beta$, where $\beta = (\beta_0, \beta_1)^\top$ with $\beta_0 = \tilde{\beta}_0 + 2\tilde{\beta}_2$ and $\beta_1 = \tilde{\beta}_1$. What is $X$?

(c) According to the previous item we have $Y = X\beta + e$. Why can we compute the ordinary least squares estimator $\hat{\beta}$ of $\beta$?

(d) Suppose that $\theta = \beta_0 - \beta_1$. An obvious estimator of $\theta$ is $\hat{\theta} = \hat{\beta}_0 - \hat{\beta}_1$. Is it an unbiased estimator of $\theta$?

(e) Show that $\sigma^2_{\hat{\theta}} := \mathbb{V}\mathrm{ar}\,\hat{\theta} = \sigma^2(1, -1)(X^\top X)^{-1}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{\sigma^2}{n}\frac{\sum_i (x_i+1)^2}{\sum_i (x_i-\bar{x})^2}$.

(f) What is the distribution of $\hat{\theta}$?

(g) We use $\widehat{\sigma^2} = \frac{\hat{e}^\top \hat{e}}{n-2}$ to estimate $\sigma^2$, where $\hat{e}$ is the vector of residuals. Let $\tau$ be the statistic

$$\tau = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}},$$

with $s_{\hat{\theta}} = \sqrt{\frac{\widehat{\sigma^2}}{n}\frac{\sum_i (x_i+1)^2}{\sum_i (x_i-\bar{x})^2}}$. What is the distribution of $\tau$? Give a very brief (rough) explanation of your answer, but no detailed computations.

(h) Use the distribution of $\tau$ to construct a $(1-\alpha)$-confidence interval for $\theta$.

(i) After performing an experiment with $n = 62$ data points, one obtains the values $\hat{\theta} = 0.60$, $s_{\hat{\theta}} = 0.40$. Choose $\alpha = 0.05$ and give a numerical confidence interval for $\theta$ based on $\tau$.

(j) One is interested in testing the null hypothesis $H_0 : \beta_0 = \beta_1$ against the alternative $\beta_0 \neq \beta_1$. Should the null hypothesis be rejected for the experiment of the previous item?

37. Let $X_1, \ldots, X_n$ be independent random variables with a common exponential distribution having parameter $\lambda > 0$. Let $X = \sum_{i=1}^n X_i$. We know that $X$ has a Gamma distribution.

   (a) Let $n > k \geq 0$. Show that $\mathbb{E}\,\frac{1}{X^k} = \frac{\lambda^k}{(n-1)\cdots(n-k)}$.

   (b) Write down the joint density of $(X_1, \ldots, X_n)$ and show that the maximum likelihood estimator of $\lambda$ is given by $\hat{\lambda} = \frac{n}{X}$.

   (c) Compute $\mathbb{E}\,\hat{\lambda}$, $\mathbb{E}\,\hat{\lambda}^2$ and show that $\mathbb{V}\mathrm{ar}\,\hat{\lambda} = \frac{n^2\lambda^2}{(n-1)^2(n-2)}$.

   (d) Show by using the previous item or by a direct computation that the mean squared error of $\hat{\lambda}$ equals $\frac{\lambda^2(n+2)}{(n-1)(n-2)}$.

   (e) Verify that with $\bar{X} = X/n$

   $$\sqrt{n}\left(\frac{\hat{\lambda}}{\lambda} - 1\right) = -\frac{1}{\lambda\bar{X}}\lambda\sqrt{n}\left(\bar{X} - \frac{1}{\lambda}\right)$$

   and deduce from the Central Limit Theorem for the sample mean that $\sqrt{n}(\frac{\hat{\lambda}}{\lambda} - 1)$ has $N(0,1)$ as limit distribution.

   (f) Compute the Fisher information $I(\lambda) = \frac{1}{\lambda^2}$ and verify that the correct answer to the previous question is in agreement with a general result.

   (g) Compute a $(1-\alpha)$-confidence interval of the type $(\ell_1(\bar{X}), r_1(\bar{X}))$ for $\lambda$ based on the asymptotic distribution of the maximum likelihood estimator.

   (h) Compute a $(1-\alpha)$-confidence interval of the type $(\ell_2(\bar{X}), r_2(\bar{X}))$ for $\lambda$ based on the asymptotic distribution of the sample mean and show that $\ell_1(\bar{X}) - \ell_2(\bar{X})$ tends to zero for $n \to \infty$.

38. Let $X_1, X_2, \ldots$ be a sequence of independent random variables, all having a Poisson distribution with parameter $\lambda > 0$. Put $S_0 = 0$, $S_n = \sum_{j=1}^n X_j$ for $n \geq 1$ and define $T = \min\{n \geq 1 : S_n > 0\}$, the first moment $S_n$ gets positive. Note that $T = k$ is equivalent to $S_k > 0$ and $S_{k-1} = 0$. Furthermore we have independent random variables $T_1, \ldots, T_n$, all having the same distribution as $T$. To express the dependence on the parameter $\lambda$, we write $\mathbb{P}_\lambda$ etc.

   (a) Show that $\mathbb{P}_\lambda(T = k) = (1 - e^{-\lambda})e^{-\lambda(k-1)}$ for $k \geq 1$.

   (b) We observe that $T$ has a familiar distribution. What are $\mathbb{E}\,T$ and $\mathbb{V}\mathrm{ar}\,T$?

Assume that the $T_1, \ldots, T_n$ are observed. Let $p_\lambda(t_1, \ldots, t_n) = \mathbb{P}_\lambda(T_1 = t_1, \ldots, T_n = t_n)$ and $L(\lambda | T_1, \ldots, T_n) = p_\lambda(T_1, \ldots, T_n)$. Let $\lambda_1 > \lambda_0 > 0$.

(c) Compute the *likelihood ratio* $\Lambda := \frac{L(\lambda_0|T_1,\ldots,T_n)}{L(\lambda_1|T_1,\ldots,T_n)}$. It is convenient to express $\Lambda$ in terms of $p_0 = 1 - e^{-\lambda_0}$ and $p_1 = 1 - e^{-\lambda_1}$.

(d) The Neyman-Pearson test for testing $H_0 : \lambda = \lambda_0$ against $H_A : \lambda = \lambda_1$ at significance level $\alpha$ rejects for small values of $\Lambda$. Show that this test is equivalent to rejecting for small values of $\sum_{j=1}^n T_j$, say $\sum_{j=1}^n T_j < c(\alpha)$.

(e) Show by invoking the Central Limit Theorem that approximately $c(\alpha) = \frac{n - z(\alpha)\sqrt{n(1-p_0)}}{p_0}$, where as usual $z(\alpha)$ is such that $\Phi(-z(\alpha)) = \alpha$.

(f) Is the Neyman-Pearson test uniformly most powerful for testing $H_0 = \lambda_0$ against the alternative $H_A : \lambda > \lambda_0$?

(g) Compute the (asymptotic) power of the test at $\lambda_1$ in terms of an expression involving the cumulative distribution of the standard normal distribution and show that it converges to 1 as $n \to \infty$.

39. Consider the model for multiple regression, $Y = X\beta + e$. Assume that $Y$ and $e$ are $n$-dimensional random vectors, $\beta = (\beta_0, \beta_1)^\top \in \mathbb{R}^2$ and $X \in \mathbb{R}^{n\times 2}$. We split the observations $Y$ in a vector $Y^1$ of length $k \le n$ and a vector $Y^2$ of length $m = n - k$, $Y^1 = (Y_1, \ldots, Y_k)^\top$ and $Y^2 = (Y_{k+1}, \ldots, Y_n)^\top$. Correspondingly, we split the design matrix $X$, which has the following special form

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_k & \mathbf{1}_k \\ \mathbf{1}_m & \mathbf{0}_m \end{pmatrix}.$$

Here $\mathbf{1}_k$ ($\mathbf{1}_m$) is a $k$-dimensional ($m$-dimensional) vector whose elements are all equal to 1, $\mathbf{0}_m$ is an $m$-dimensional vector whose elements are all equal to 0. The experiment thus uses the dummy variables 1 and 0 to indicate whether a population item possesses a certain property or not. Think of people, where females get the value 1, whereas males are 'worthless'.

(a) Show that $X^\top X = \begin{pmatrix} k+m & k \\ k & k \end{pmatrix}$, and compute its inverse.

(b) Show that the Least Squares estimator $\hat\beta$ of $\beta$ is given by

$$\hat\beta = \begin{pmatrix} \hat\beta_0 \\ \hat\beta_1 \end{pmatrix} = \begin{pmatrix} \overline{Y^2} \\ \overline{Y^1} - \overline{Y^2} \end{pmatrix},$$

where $\overline{Y^1} = \frac{1}{k}\sum_{i=1}^k Y_i$ and $\overline{Y^2} = \frac{1}{m}\sum_{j=k+1}^n Y_j$.

(c) Compute expectation and variance of $\hat\beta_1$.

(d) The designer of the experiment is able to control the sizes $k$ and $m$ of the sub-populations. How should she choose $k$ and $m$ such that the mean squared error of $\beta_1$ is minimal? *For convenience you may assume that $n$ is even.*

Assume that the error term $e$ has a multivariate normal distribution with zero expectation and covariance matrix equal to $\sigma^2 I$, with $\sigma \in (0, \infty)$.

(e) Determine the distribution of

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma\sqrt{\frac{1}{k} + \frac{1}{m}}}.$$

(f) Let $S^2$ be the 'usual' estimator of $\sigma^2$ and its positive root $S$. Show that $S^2 = (\sum_{i=1}^{k}(Y_i - \overline{Y1})^2 + \sum_{j=k+1}^{n}(Y_j - \overline{Y2})^2)/(n-2)$ and characterize the distribution of $S^2$.

(g) Derive from general results the distribution of

$$\frac{\hat{\beta}_1 - \beta_1}{S\sqrt{\frac{1}{k} + \frac{1}{m}}}.$$

(h) We test at significance level $\alpha$ the hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_A : \beta_1 \neq 0$. The total sample size is 100, $k = 50$, $\alpha = 0.01$. Numerical results yield the values $\hat{\beta}_1 = -5.00$ and $S = 12.50$. Construct a confidence interval for $\beta_1$.

(i) Think of the responses $Y_i$ as income and a population of women and men as at the beginning of the exercise. Do women significantly earn (1) more or (2) less then men, or is there (3) no significant difference?

40. Let Let $X_1, \ldots, X_n$ be independent random variables with a common *uniform* distribution on an interval $[0, \theta]$, where $\theta > 0$ is an unknown parameter. It is known that the maximum likelihood estimator of $\theta$ is given by $X_{(n)} := \max\{X_1, \ldots, X_n\}$. Obviously, we have $X_{(n)} \leq \theta$.

(a) Show that $X_{(n)}$ has distribution function given by $F_{X_{(n)}}(x) = (\frac{x}{\theta})^n$ for $x \in [0, \theta]$.

(b) Show that $X_{(n)}$ is a consistent estimator of $\theta$: since $X_{(n)} \leq \theta$, it is sufficient to show that $\mathbb{P}(X_{(n)} < \theta - \delta) \to 0$ as $n \to \infty$ for all $\delta > 0$.

(c) Compute the density of $X_{(n)}$ and show that $\mathbb{E}\, X_{(n)} = \frac{n}{n+1}\theta$.

(d) For which constant $c_n$ is $\hat{\theta}_n := c_n X_{(n)}$ an unbiased estimator?

(e) One can show that $\mathbb{E}\, X_{(n)}^2 = \frac{n}{n+2}\theta^2$. Show that the variance of $X_{(n)}$ equals $\frac{n}{(n+1)^2(n+2)}\theta^2$ and from this that the mean squared error of $X_{(n)}$ equals $\frac{2}{(n+2)(n+1)}\theta^2$.

(f) Deduce from the previous item that the variance of $\hat{\theta}_n$ is equal to $\frac{\theta^2}{n(n+2)}$.

(g) Which of the estimators $X_{(n)}$ and $\hat{\theta}_n$ is preferred and why?

(h) Let $Y_n = n(\theta - X_{(n)}) \geq 0$. Show that the distribution function $F_{Y_n}$ of $Y_n$ is given by $F_{Y_n}(y) = 1 - (1 - \frac{y}{n\theta})^n$ for $y > 0$. Show that $Y_n$ converges in distribution to $Y$, where $Y$ has an exponential distribution with mean $\theta$.

(i) Let $W_n = n(\theta - \hat{\theta}_n)$. Show that $W_n = Y_n - X_{(n)}$ and that $W_n$ converges in distribution to $W := Y - \theta$ (give a precise argument).

(j) Compute $\mathbb{E}\, Y^2$ and $\mathbb{E}\, W^2$. Are the results in agreement with those of items (e) and (f).

(k) All mean squared errors and variances above tend to zero by a factor roughly proportional to $\frac{1}{n^2}$. This is in contrast with the usual behavior of maximum likelihood estimators and the Cramér-Rao bound. Part of the explanation is that the the Fisher information $I(\theta)$ (in a single observation) is not well defined. Why is this the case? *It may help to sketch the density of $X_1$ as a function of $\theta$ for $\theta > 0$ (the likelihood)* .

41. Here we consider two *independent* samples, one from a $N(\nu, \sigma^2)$ distribution, one from a $N(\mu, \sigma^2)$ distribution. Note that the variances of the two distributions are the same. Specifically, we have *iid* random variables $X_1, \ldots, X_n$ with a common $N(\mu, \sigma^2)$ distribution, and *iid* random variables $Y_1, \ldots, Y_m$ with a common $N(\nu, \sigma^2)$ distribution. All $X_i$ are also independent of all $Y_j$. It is of interest whether or not $\mu = \nu$. Let $\delta = \mu - \nu$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $\overline{Y} = \frac{1}{m}\sum_{j=1}^m Y_j$. Assume that $\sigma^2$ is known.

(a) Show that $\overline{X} - \overline{Y}$ has a $N(\delta, (\frac{1}{n} + \frac{1}{m})\sigma^2)$ distribution.

(b) Give a $(1 - \alpha)$-confidence interval for $\delta$.

(c) Write the joint density of $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$. Note that the unknown parameters are $\mu$ and $\nu$.

(d) Show that the maximum likelihood estimators of $\mu$ and $\nu$ are given by $\hat{\mu} = \overline{X}$ and $\hat{\nu} = \overline{Y}$.

(e) Suppose that it is known that $\nu = \mu$. In this case the likelihood only depends on the unknown parameter $\mu$. Show that in this situation the maximum likelihood estimator is given by $\tilde{\mu} = \frac{n\overline{X}+m\overline{Y}}{n+m}$.

(f) Let $S_X^2 = \sum_{i=1}^n (X_i - \overline{X})^2$ and $S_Y^2 = \sum_{j=1}^m (Y_j - \overline{Y})^2$. Use

$$\tilde{\mu} = \overline{X} + \frac{m}{m+n}(\overline{Y} - \overline{X}) = \overline{Y} + \frac{n}{m+n}(\overline{X} - \overline{Y})$$

to write

$$\sum_{i=1}^n (X_i - \tilde{\mu})^2 + \sum_{j=1}^m (Y_j - \tilde{\mu})^2 = S_X^2 + S_Y^2 + \frac{nm}{n+m}(\overline{X} - \overline{Y})^2.$$

(g) Let $H_0 : \mu = \nu$ and $H_A : \mu \neq \nu$. Show that the (generalized) likelihood ratio test statistic is equal to

$$\Lambda = \exp\big( -\frac{1}{2\sigma^2}\frac{nm}{n+m}(\overline{X} - \overline{Y})^2 \big).$$

(h) Suppose that $H_0$ is tested against $H_A$ at significance level $\alpha$ using $\Lambda$ as the test statistic. Describe precisely the critical region of the test (then $H_0$ is rejected) in terms of $\overline{X} - \overline{Y}$.

(i) Give a connection between the answers to (a) and (h).

42. Consider the linear regression model $Y = X\beta + e$, where $Y$ and $e$ are $2n$-dimensional random vectors, $X \in \mathbb{R}^{2n\times 2}$ and $\beta = (\beta_0, \beta_1)^\top$. Moreover, the design matrix $X$ is of the special form

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n & \mathbf{1}_n \\ \mathbf{1}_n & -\mathbf{1}_n \end{pmatrix},$$

where $\mathbf{1}_n$ is an $n$-dimensional vector whose elements are all equal to 1. We furthermore assume that the elements of the vector $e$ are *iid* random variables with zero mean and variance $\sigma^2$. Let $\overline{Y}_1 = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\overline{Y}_2 = \frac{1}{n}\sum_{j=n+1}^{2n} Y_j$.

(a) Show that the least squares estimator of $\beta$ is given by

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(\overline{Y}_1 + \overline{Y}_2) \\ \frac{1}{2}(\overline{Y}_1 - \overline{Y}_2) \end{pmatrix}.$$

What is the variance of $\hat{\beta}_1$?

(b) Let $\hat{e}$ be the vector of residuals, $\hat{e} = Y - X\hat{\beta}$. Show that $\hat{e}^\top \hat{e} = \sum_{i=1}^{n}(Y_i - \overline{Y}_1)^2 + \sum_{j=n+1}^{2n}(Y_j - \overline{Y}_2)^2$.

(c) What is $\mathbb{E}\,(\hat{e}^\top \hat{e})$? Give an unbiased estimator $\widehat{\sigma^2}$ of $\sigma^2$.

A group of $2n$ people is selected for a screen test to find out whether they are suitable as a candidate to present a new television programme. The first $n$ persons are female and the last $n$ persons are male. Since beauty is thought of as a potential criterion for selection, the second column of the design matrix is chosen to reflect this phenomenon. If somebody passes the test, the corresponding $Y$-value is 1 and 0 otherwise. Somebody is interested to see whether gender is of influence in the selection of candidates in the sense that women have a better chance to pass the test.

(d) Explain that the error term $e$ cannot have a (multivariate) normal distribution.

(e) Formulate a hypothesis testing problem in terms of the parameter $\beta$ that reflects the research issue. Give a suitable test statistic and describe the rejection region.

Although the usual normality assumptions are not valid, we will ignore this and assume that, as usual, relevant test statistics have a $t$-distribution with the appropriate number of degrees of freedom (this can be justified as an approximation if $n$ is not too small). Suppose that $n = 10$ and that 6 women and 4 men pass the test.

(f) What is the result of the test at significance level $\alpha = 0.05$? Can we conclude from the observations that women have an advantage to pass the test? And what is the conclusion if $\alpha = 0.01$?

43. Let $X_1, \ldots, X_n$ be independent random variables with a distribution having a (continuous) density $f(x;\lambda)$ such that $f(x;\lambda) = 2\lambda x \exp(-\lambda x^2)$ for $x \geq 0$.

(a) Show that $f(x;\lambda) = 0$ for $x < 0$. *Hint:* computing an integral helps.

(b) Show that $\mathbb{E}\,X_1 = \frac{1}{2}\sqrt{\frac{\pi}{\lambda}}$. *Hint:* write the expectation as an integral $\int \ldots \mathrm{d}x$, make the change of variable $y = \sqrt{2\lambda}x$ and recognize, up to a constant, a known variance.

(c) Show that $\mathbb{E}\,X_1^2 = \frac{1}{\lambda}$. *Hint:* use integration by parts.

(d) Show that $\mathbb{E}\,X_1^4 = \frac{2}{\lambda}\mathbb{E}\,X_1^2$.

(e) Let $\ell(\lambda; X_1)$ be the log-likelihood, when $X_1$ is observed. Show that $\dot{\ell}(\lambda; X_1) = \frac{1}{\lambda} - X_1^2$. Compute $\mathbb{V}\mathrm{ar}\,\dot{\ell}(\lambda; X_1)$.

(f) Compute, using the results of the previous items, the Fisher information $I(\lambda|X_1) = I(\lambda)$, if only $X_1$ is observed.

(g) Compute the same Fisher information $I(\lambda)$ above by a different method.

(h) Compute the maximum likelihood estimator $\hat{\lambda}_n$ of $\lambda$ for the case where the full sample $X_1, \ldots, X_n$ is observed.

(i) Characterize the asymptotic distribution of $\hat{\lambda}_n$ for $n \to \infty$?

(j) What is the maximum likelihood estimator $\hat{\theta}_n$ of $\theta$, for $\theta = \frac{1}{\lambda}$ if $X_1, \ldots, X_n$ is observed?

(k) Is $\hat{\theta}_n$ an unbiased estimator of $\theta$?

(l) Compute the mean squared error of $\hat{\theta}_n$.

(m) Show that $\hat{\theta}_n$ is a consistent estimator of $\theta$.

(n) Use the Central Limit Theorem to find the asymptotic distribution $(n \to \infty)$ of $\hat{\theta}$.

44. Here are two statements. (1) The popularity of political parties is measured in terms of their number of members. (2) The popularity of political parties is measured in terms of their number of votes at the last held elections. Both statements contain some truth, and one may wonder whether the number of members of parties can be used to predict the number of votes in the first coming elections. We will test this hypothesis in the framework of linear regression. In technical terms, the $x_i$ will be the membership sizes and $Y_i$ the number of votes, both measured for The Netherlands at the election date September 12, 2012. In the table below we have the figures for the 11 parties that ended up with representatives (there are 150 seats in total in the parliament) in the parliament after September 12, 2012.

| party | members | seats |
|---|---:|---:|
| CDA | 61294 | 13 |
| CU | 24701 | 5 |
| D66 | 21985 | 12 |
| GL | 26505 | 4 |
| PvdA | 54279 | 38 |
| PvdD | 12250 | 3 |
| PVV | 1 | 15 |
| SGP | 28048 | 3 |
| SP | 44186 | 15 |
| VVD | 38412 | 41 |
| 50+ | 1321 | 2 |

There is one strange party in the sense that it has one member only, the membership is not open to anybody else than the party leader.

(a) One of the assumptions used to apply the usual regression model is that the $Y_i$ are independent. Is this assumption satisfied for the 11 parties in the sample?

(b) Because of this one less democratic party (in the sense of absence of internal democracy), we omit this party from this and the remaining questions. How would you answer the previous question for the 10 remaining parties.

(c) In the standard regression model (whose validity we assume in what follows) there are the parameters $\beta_0$ and $\beta_1$. How would you describe in terms of these parameters the null hypothesis that the number of members has no predictive power. How would you describe the alternative? (I see two possibilities from which you can choose, and your choice should be consistent with your answers to the remaining questions.)

(d) Compute the least squares estimators of $\beta_0$, $\beta_1$ from the data. You may use the rounded numbers $\sum x_i = 313000$, $\sum x_i^2 = 129000000$, $\sum x_i y_i = 5714000$, the sum of squared residuals is 1184 (simplify further in your calculations if you don't have a calculator).

(e) Give an estimate of $\sigma^2$, the assumed common variance of the $Y_i$.

(f) Test the null hypothesis against the alternative as you have formulated it when answering question 44c by using a rejection region and a suitable test statistic $T$.

(g) Compute the $p$-value for the chosen test statistic $T$.

(h) Give a one- or two-sided confidence interval for $\beta^1$, depending on you choise of the alternative hypothesis.

(i) Check whether the estimate of $\beta_1$ is in the confidence interval. Is this in agreement with the result of the test?

45. Let $X$ be a random variable whose distribution function is determined by the formula $F(x) = 1 - x^{-\lambda}$, where $\lambda$ is a positive parameter.

(a) The given formula cannot be correct for all $x \in \mathbb{R}$. What are the values of $x$ for which the formula makes sense? Make a sketch of the graph of $F$ for your favourite value of $\lambda$ and compute the density of $X$.

(b) Let $k$ be a positive integer and assume that $\lambda > k$. Show that $\mathbb{E}\, X^k = \frac{\lambda}{\lambda - k}$.

(c) What is the variance of $X$, if it is finite?

In the remainder of this exercise we are dealing with a sample $X_1, \ldots, X_n$ of independent, identically distributed random variables with common distribution function $F$. By $\overline{X}$ we denote the sample average.

(d) Show that the moment estimator $\hat{\lambda}_{\text{mom}}$ of $\lambda$ is given by $\frac{\overline{X}}{\overline{X}-1}$

(e) Show that $\hat{\lambda}_{\text{mom}}$ is a consistent estimator of $\lambda$ (Note that $\hat{\lambda}_{\text{mom}}$ is of the type $g(\overline{X})$ and use the law of large numbers).

(f) Show that

$$\sqrt{n}(\hat{\lambda}_{\text{mom}} - \lambda) = -\frac{\lambda - 1}{\overline{X} - 1}\sqrt{n}(\overline{X} - \frac{\lambda}{\lambda - 1}).$$

(g) Assume $\lambda > 2$. Show, use the central limit theorem and additional arguments, that $\sqrt{n}(\hat{\lambda}_{\text{mom}} - \lambda)$ converges in distribution to a random variable having a normal distribution with variance $\frac{\lambda(\lambda-1)^2}{\lambda-2}$.

(h) Compute the maximum likelihood estimator $\hat{\lambda}_{\text{MLE}}$ of $\lambda$.

(i) Compute the Fisher information $I(\lambda)$ (for $n = 1$).

(j) What is the limit distribution of $\sqrt{n}(\hat{\lambda}_{\text{MLE}} - \lambda)$?

(k) Which of the two estimators $\hat{\lambda}_{\text{mom}}$ and $\hat{\lambda}_{\text{MLE}}$ should be preferred?

46. A random variable $X$ is said to have a *log-normal distribution* with parameters $\mu$ and $\sigma^2$ if $X = \exp(Y)$, where $Y$ has a $N(\mu, \sigma^2)$ distribution and $\sigma > 0$. An alternative representation of such an $X$ is $X = \exp(\mu + \sigma Z)$, where $Z$ has a standard normal distribution. Below we will consider a sample $X_1, \ldots, X_n$ of independent $X_i$, all having the same log-normal distribution. The parameter $\sigma$ is considered to be known.

(a) Show that the density of the log-normal distribution is

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\big(-\frac{1}{2\sigma^2}(\log x - \mu)^2\big).$$

(b) Let $\hat{\mu}$ be the maximum likelihood estimator of $\mu$. Show that $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n \log X_i$.

We consider the testing problem $H_0 : \mu = \mu_0$ against $H_A : \mu = \mu_A$, where $\mu_A > \mu_0$, at significance level $\alpha$.

(c) Show that the most powerful test rejects the null-hypothesis for ‘large values’ of $\hat{\mu}$ by manipulating the likelihood ratio.

(d) To make the previous item more precise, show that $H_0$ is rejected when $\hat{\mu} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$, where $z_\alpha$ is such that $\Phi(z_\alpha) = 1 - \alpha$. *Hint: use that* $X_i = \exp(\mu_0 + \sigma Z_i)$ *with* $Z_i$ *standard normal under* $H_0$.

(e) Is the test also uniformly most powerful for the testing problem $H_0 : \mu = \mu_0$ against $H_A : \mu > \mu_0$? Same question for the testing problem $H_0 : \mu = \mu_0$ against $H_A : \mu < \mu_0$.

(f) Show that the power of the test of item (d) in $\mu = \mu_A$ is equal to $1 - \Phi\big((\mu_0 - \mu_A)\frac{\sqrt{n}}{\sigma} + z_\alpha\big)$. What happens to the power if the sample increasing to infinity? Is this desirable?

47. Consider the standard regression model $Y = X\beta + e$, where $Y$ is an $n$-dimensional random vector, $X$ a $n \times p$ matrix, $\beta = (\beta_0, \dots, \beta_{p-1})^\top$, the parameter vector, $p$-dimensional and $e$ $n$-dimensional. Also the usual independence and normality assumptions are satisfied. Assume $X$ has rank $p$, write $P = X(X^\top X)^{-1}X^\top$ and let $\hat{\beta}$ be the usual least squares estimator.

We introduce an additional regressor $\xi \in \mathbb{R}^n$, which is such that $g := \xi^\top(I - P)\xi = \xi^\top(I - X(X^\top X)^{-1}X^\top)\xi > 0$. Here is some additional notation and results.

$$
\begin{aligned}
\tilde{X} &:= \begin{pmatrix} X & \xi \end{pmatrix} \in \mathbb{R}^{n\times(p+1)} \\
(\tilde{X}^\top\tilde{X})^{-1} &= \begin{pmatrix} (X^\top X)^{-1} + \frac{1}{g}(X^\top X)^{-1}X^\top\xi\xi^\top X(X^\top X)^{-1} & -\frac{1}{g}(X^\top X)^{-1}X^\top\xi \\ -\frac{1}{g}\xi^\top X(X^\top X)^{-1} & \frac{1}{g} \end{pmatrix} \\
\tilde{P} &:= \tilde{X}(\tilde{X}^\top\tilde{X})^{-1}\tilde{X}^\top = P + \frac{1}{g}(I - P)\xi\xi^\top(I - P).
\end{aligned}
$$

As an alternative to the given standard model, one may also consider the extended model $Y = \tilde{X}\beta' + e$, where $\beta'$ now becomes $(p+1)$-dimensional, $\beta' = (\beta_0, \dots, \beta_p)^\top$, and the rows of $\tilde{X}$ now consist of $p+1$ regressors. This extended model should give a better fit than the original one. To measure the fit of the original model we look at $\hat{e} = Y - \hat{Y}$, where $\hat{Y} = X\hat{\beta}$ and for the extended model one considers $\tilde{e} = Y - \tilde{Y}$ with $\tilde{Y} = \tilde{X}\tilde{\beta}$ and $\tilde{\beta}$ the $(p+1)$-dimensional least squares estimator. By $||v||$ we denote the usual norm of a vector $v$, $||v|| = \sqrt{v^\top v}$.

(a) Show that the extended model gives a better fit, $||\tilde{e}|| \le ||\hat{e}||$, by computing

$$||\tilde{e}||^2 = ||\hat{e}||^2 - \frac{1}{g}\big(Y^\top(I - P)\xi\big)^2.$$

Show also that $\mathbb{E}\,\frac{1}{g}\big(Y^\top(I - P)\xi\big)^2 = \sigma^2$ and hence that $\mathbb{E}\,||\tilde{e}||^2 = (n - p - 1)\sigma^2$.

(b) Show for the extended model that

$$\tilde{\beta} = \begin{pmatrix} \hat{\beta} - (X^\top X)^{-1} X^\top \xi \, \tilde{\beta}_p \\ \frac{1}{g} \xi^\top (Y - X\hat{\beta}) \end{pmatrix}, \tag{2}$$

where $\tilde{\beta}_p$ is the last element of the vector $\tilde{\beta}$.

(c) Show directly from (2) that $\tilde{\beta}_p$ is an unbiased estimator of $\beta_p$. (Note that $\tilde{\beta}_p = \frac{1}{g}\xi^\top (I - P)Y$ and use the extended model.)

(d) Show directly from (2) that $\mathbb{V}\mathrm{ar}\,(\tilde{\beta}_p) = \frac{\sigma^2}{g}$. How could you have already known this from the theory?

Here are some figures. Suppose in an experiment with $n = 23$ and $p = 2$, one finds $g = 0.81$, $\xi^\top (I - P)Y = 0.324$, $\hat{e}^\top \hat{e} = 1.1421$, $\tilde{e}^\top \tilde{e} = 1.0125$; these figures should allow for 'easy' calculations.

(e) Construct a numerical 95%-confidence interval for $\beta_p$ on the basis of $\tilde{\beta}_p$ with the given data.

(f) Consider the hypotheses $H_0 : \beta_p = 0$ and $H_A : \beta_p \neq 0$. Will the null hypothesis be rejected at the significance level $\alpha = 0.05$ using $\tilde{\beta}_p$ as a test statistic? Same question for $\alpha = 0.01$.

48. Let $X$ have an exponential distribution with parameter $\lambda > 0$, so $X$ has density $f(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$. Let $U$ be *independent* of $X$ such that $\mathbb{P}(U = +1) = \mathbb{P}(U = -1) = \frac{1}{2}$, and put $Z = U\sqrt{X}$.

(a) We are interested in the distribution function $F_Z$ of $Z$. Show, split the event $\{Z \leq z\}$ into two sub-events according to $U = \pm 1$, that

$$F_Z(z) = \begin{cases} \frac{1}{2}(1 - \exp(-\lambda z^2)) + \frac{1}{2} & \text{if } z \geq 0 \\ \frac{1}{2}\exp(-\lambda z^2) & \text{if } z < 0. \end{cases}$$

(b) Show that the density $f_Z$ of $Z$ is given by $f_Z(z) = \lambda |z| \exp(-\lambda z^2)$.

(c) Give a rough (but not too rough) sketch of the graph of $f_Z$. Thereby you pay attention to the values of $f_Z$ for $z$ near zero and for $z \to \pm\infty$.

(d) From the previous item you can immediately deduce what $\mathbb{E}\, Z$ is. How? Verify your answer by using the definition of $Z$.

(e) Show that $\mathbb{V}\mathrm{ar}\,(Z^2) = \frac{1}{\lambda^2}$. *Hint: use the definition of $Z$.*

In the remainder of this exercise we assume to have an IID sample $Z_1, \ldots, Z_n$, each of the $Z_i$ having the density as in (b), which constitute our observations.

(f) Give the expression for the joint density of the vector $(Z_1, \ldots, Z_n)$ and show that the maximum likelihood estimator $\hat{\lambda}$ of $\lambda$ is given by $\hat{\lambda} = n / \sum_{i=1}^n Z_i^2$.

(g) Compute the Fisher information $I(\lambda)$ in a single observation $Z_1$.

(h) Give an approximate $(1-\alpha)$-confidence interval for $\lambda$ based on $\hat{\lambda}$. (Note that the limits of this interval should not depend on the unknown $\lambda$.)

(i) Suppose the distribution of $Z$ is reparametrized by using $\tau = 1/\lambda$. What is the maximum likelihood estimator of $\tau$? Show that it is unbiased.

(j) Many of the answers should be familiar to you. The explanation is as follows. Although $Z \neq X$, knowing $Z$ also tells you what $X$ is (and something similar holds for the $Z_i$). Why?

49. Consider a random variable $X$ with density function $f$ given by $f(x) = \frac{1}{2}\lambda^2 \exp(-\lambda\sqrt{x})$ for $x \geq 0$ (and zero otherwise), where $\lambda$ is a positive parameter. You may safely assume that $f$ is indeed a density.

(a) Show that $\mathbb{E}\, X = \frac{6}{\lambda^2}$. *Hint: write down the integral, make the substitution $x = u^2$ and be clever from that point on.*

In the remainder of this exercise you will also need $\mathbb{E}\sqrt{X} = \frac{2}{\lambda}$. We will be interested in the hypothesis testing problem $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda = \lambda_1$ at some significance level $\alpha$. The observations are $X_1, \ldots, X_n$, IID under each of the hypotheses. Our test statistic will be the Likelihood Ratio, we call it LR.

(b) Write down in terms of the observations the formula for LR.

(c) The Likelihood Ratio test rejects $H_0$ for small values of LR. Show that this test is equivalent to rejecting $H_0$

i. for small values of $T_n := \sum_{i=1}^n \sqrt{X_i}$ if $\lambda_0 < \lambda_1$, and

ii. for big values of $T_n$ if $\lambda_0 > \lambda_1$.

Henceforth we will be interested in the case $\lambda_0 > \lambda_1$. Although it is possible to determine the distribution of $T_n$ under $H_0$, we use the Central Limit Theorem to approximate this distribution. Let $Q_n = \frac{\lambda_0}{\sqrt{2n}}(T_n - \frac{2n}{\lambda_0})$.

(d) Show that, under the null hypothesis, $Q_n$ converges in distribution to a random variable having the standard normal distribution.

(e) Show that the Likelihood Ratio test rejects $H_0$ if (approximately) $T_n > \frac{2n}{\lambda_0} + z_\alpha \frac{\sqrt{2n}}{\lambda_0}$.

(f) Suppose that in a practical experiment $n = 200$, $\lambda_0 = 1$ and that $T_n = 750$ is observed. Compute the $p$-value. If $\alpha = 0.05$, should one reject $H_0$?

(g) Is the Likelihood Ratio uniformly most powerful for the testing problem $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda < \lambda_0$ at the same significance level $\alpha$?

(h) Suppose we change the testing problem in the previous item into $H_0 : \lambda \geq \lambda_0$ versus $H_A : \lambda < \lambda_0$ and that one decides to reject $H_0$ if (again) $T_n > \frac{2n}{\lambda_0} + z_\alpha \frac{\sqrt{2n}}{\lambda_0}$. Show that the significance level of this test is again equal to $\alpha$. *Hint: Recall that according to the significance level with a composite hypothesis one has to compute* $\sup_{\lambda \geq \lambda_0} \mathbb{P}_\lambda(\text{Reject } H_0)$. *You use again a normal approximation, which now depends on* $\lambda$, *to first compute* $\mathbb{P}_\lambda(\text{Reject } H_0)$ *and show that this probability is decreasing in* $\lambda$.

50. In this exercise we relate *popularity*, or *audience ratings* to *market shares* of television programmes. Popularity ratings of television programmes are the numbers of viewers reported as percentages of the *total* population, whereas market shares of programmes are reported as percentages of that part of the population that is watching TV at the same time. The underlying figures in Table 1 list the results of ten popular programmes of the public nets in The Netherlands on December 5, 2014. The time slots refer to different periods over the day, which explains that the sum of the first row largely exceeds 100%. In the table 'ms' denotes market share and 'pr' popularity rating.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ms | 31.9 | 28.8 | 27.6 | 30.9 | 20.2 | 26.5 | 20.3 | 23.1 | 14.0 | 23.1 |
| pr | 12.0 | 12.0 | 10.7 | 8.8 | 8.4 | 8.2 | 6.7 | 5.8 | 5.6 | 5.6 |

Table 1: the data

In a regression model we take 'pr' as the explanatory variable (the $x_i$) and 'ms' as the response (the $y_i$). The usual independence and normality assumptions are assumed to be satisfied. To facilitate the computations, we give in Table 2 some summary statistics.

$$\sum x_i = 83.80$$
$$\sum y_i = 246.40$$
$$\sum x_i y_i = 2160.37$$
$$\sum x_i^2 = 758.98$$
$$\sum y_i^2 = 6349.22$$
$$\sum (x_i - \bar{x})^2 = 56.74$$
$$\sum (y_i - \bar{y})^2 = 277.92$$
$$\sqrt{\sum (x_i - \bar{x})^2} = 7.53$$
$$\sqrt{\sum (y_i - \bar{y})^2} = 16.67$$
$$\sum (y_i - \hat{y}_i)^2 = 117.05$$

Table 2: summary statistics

(a) Compute from the data the least squares estimates $\hat{\beta}_1$ and $\hat{\beta}_0$.[1]

(b) Test the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_A : \beta_1 > 0$ at the significance level $\alpha = 0.01$.

(c) Consider the testing problem with null hypothesis $H_0 : \beta_1 = \beta_1^0$ and alternative $H_A : \beta_1 > \beta_1^0$ at a significance level $\alpha$. A *one sided* confidence interval for $\beta_1$ are those $\beta_1^0$ that are not rejected by a test for the this testing problem. Give a theoretical one sided $(1 - \alpha)$-confidence interval for $\beta_1$.

(d) Using the above data, give also a numerical one sided confidence interval for $\beta_1$ with $\alpha = 0.01$. Is your result in agreement with your answer under (b)?

51. Consider the experiment of throwing two dice (numbered 1 and 2). Let $X_i$ be the number of dots showing on dice $i$. The underlying probability model has the joint probability mass function (pmf) $\mathbb{P}(X_1 = i, X_2 = j) = \frac{1}{36}$ for all relevant values of $i$ and $j$. We consider $X := X_1 + X_2$. The pmf of $X$ is

[1] If you don't have a pocket calculator, here and in the other questions you are allowed to simplify the figures a bit for easier computations.

as given in the incomplete table below.

| $k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_k := \mathbb{P}(X = k)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | | $\frac{2}{36}$ | $\frac{1}{36}$ |

We introduce a new random variable $Y$, which is defined as the remainder of $X$ after dividing by 4. As an example, if $X = 7$, then $X = 1 \times 4 + 3$, so in this case $Y = 3$. The values of $Y$ are in the set $V_Y = \{0, 1, 2, 3\}$.

(a) Compute the missing $p_k$ from the underlying model.

(b) Compute the pmf of $Y$, so the probabilities $\mathbb{P}(Y = l)$, $l \in V_Y$.

(c) Make a sketch of the distribution function $F_Y(y)$ of $Y$ $(y \in \mathbb{R})$ and compute the jumps of $F_Y$ at the points $y = 2$ and $y = 5$. (Recall that these jumps are defined as $\Delta F_Y(y) = F_Y(y) - F_Y(y-)$.)

52. In this exercise we investigate the number of homicides in a country in relation to the density of fire arms present. In the table below (figures taken from *The Guardian*), the first column of number represents the number of homicides per 100000 inhabitants (denoted $y_i$), the second column of numbers the number of fire arms in a country per 100 inhabitants (denoted $x_i$). The figures for the United States (2.97 and 88.8) play a special role.

| | homocides ($y_i$) | fire arms ($x_i$) |
|---|---|---|
| Australia | 0.14 | 15.0 |
| Austria | 0.22 | 30.4 |
| Belgium | 0.68 | 17.2 |
| Canada | 0.51 | 30.8 |
| Denmark | 0.27 | 12.0 |
| Finland | 0.45 | 45.3 |
| Germany | 0.19 | 30.3 |
| Ireland | 0.48 | 8.60 |
| Luxembourg | 0.62 | 15.3 |
| Netherlands | 0.33 | 3.90 |
| New Zealand | 0.16 | 22.6 |
| Sweden | 0.41 | 31.6 |
| Switzerland | 0.77 | 45.7 |

For the investigation we use regression *with the US data ignored* and you may assume that all usual assumptions are in force, $Y_i = \beta_0 + \beta_1 x_i + e_i$ with $e_i \sim N(0, \sigma^2)$, etc. For actual computations you can use Table 3 with summary statistics at the end of this exercise.

(a) Show that the (estimated) regression line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is given by $\hat{y} = 0.31 + 0.0038x$ (the coefficients are rounded).

(b) What is the exact theoretical value of $\sum_i (\hat{y}_i - y_i)$, where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$?

(c) Compute the value of the usual unbiased estimator of $\sigma^2$.

(d) Compute an exact *two-sided* 90% confidence interval for $\beta_1$. Should the null hypothesis $H_0 : \beta_1 = 0$ against the *one sided* alternative $H_A : \beta_1 > 0$ be rejected at significance level 5%?

We now consider *prediction* of the response variable for a new value of the independent variable, for which the general setting is as follows. One uses observations $(x_1, Y_1), \ldots, (x_n, Y_n)$ (again the usual assumptions $Y_i = \beta_0 + \beta_1 x_i + e_i$, $e_i \sim N(0, \sigma^2)$, etc. are in force) and resulting quantities like the least squares estimators $\hat{\beta}_0$, $\hat{\beta}_1$. Let $x$ be a new value of the independent variable. Then the prediction of the response variable is denoted $\hat{Y}(x) := \hat{\beta}_0 + \hat{\beta}_1 x$, a random variable, whereas its true value is $Y(x) = \beta_0 + \beta_1 x + e$ with $e \sim N(0, \sigma^2)$ and $e$ independent of the $e_i$.

(e) Show that $\mathbb{E}\,\hat{Y}(x)$ is an unbiased estimator of $\mathbb{E}\,Y(x)$.

(f) Use the expression for the covariance matrix of $\hat{\beta}$ (here $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^\top$) to show that $\mathbb{V}\mathrm{ar}\,\hat{Y}(x) = \sigma^2 \frac{\sum_i (x_i - x)^2}{n \sum_i (x_i - \bar{x})^2}$. (In all sums, $i$ runs from 1 to $n$).

(g) The prediction error is $Y(x) - \hat{Y}(x)$. Show that its variance is equal to $\sigma^2 (1 + \frac{\sum_i (x_i - x)^2}{n \sum_i (x_i - \bar{x})^2})$.

(h) If $\sigma^2$ would be known, an exact $(1-\alpha)$ prediction interval (much like a confidence interval) for $Y(x)$ has limits $\hat{Y}(x) \pm \sigma \sqrt{1 + \frac{\sum_i (x_i - x)^2}{n \sum_i (x_i - \bar{x})^2}}\, z_{\alpha/2}$. How would you adjust this for the case where $\sigma^2$ has to be estimated to again have an exact $(1-\alpha)$ prediction interval?

(i) Compute the by the regression line predicted value number $\hat{Y}(x)$ of homicides per 100000 inhabitants for the US, with $x$ representing the number of fire arms in the US. Give also a numerical 90% prediction interval.

(j) Do you think that the estimated regression line is also valid for prediction of the US data?

53. Consider a random variable $X$ with density function $f_\theta$ given by $f_\theta(x) = \frac{1}{12}\theta^{-4} x \exp(-\sqrt{x}/\theta)$ for $x \geq 0$ (and zero otherwise), where $\theta$ is a positive parameter. You may safely assume that $f$ is indeed a density.

| $\sum x_i$ | $\sum y_i$ | $\sum x_i^2$ | $\sum x_i y_i$ | $\sum_i (\hat{y}_i - y_i)^2$ | $\sum_i (x_i - 88.8)^2$ | $\sum_i (x_i - \bar{x})^2$ |
|---|---|---|---|---|---|---|
| 308.7 | 5.23 | 9428.89 | 132.24 | 0.4774 | 57114.49 | 2098.45 |

Table 3: summary statistics

(a) Show that $\mathbb{E}\, X = 20\theta^2$. *Hint: write down the integral, make the substitution* $x = u^2\theta^2$.

(b) Show that $\mathbb{E}\sqrt{X} = 4\theta$ and $\mathbb{V}\mathrm{ar}\sqrt{X} = 4\theta^2$.

(c) Show that the Fisher information (in one observation) $I(\theta)$ is given by $I(\theta) = 4/\theta^2$.

In the remainder of this exercise we have observations $X = (X_1, \ldots, X_n)$, IID with common density $f_\theta$ as above. The parameter $\theta$ is unknown and has to be estimated.

(d) Write down in terms of the observations the likelihood $L(\theta|X)$ and the log-likelihood $\ell(\theta|X)$.

(e) Show that the maximum likelihood estimator of $\theta$ is $\hat{\theta}_n = \frac{1}{4n}\sum_{i=1}^n \sqrt{X_i}$ and that $\hat{\theta}_n$ is an unbiased estimator of $\theta$.

(f) Show that, given the observations $X_1, \ldots, X_n$, $\hat{\theta}_n$ is the best (in the sense of smallest mean squared error) unbiased estimator of $\theta$.

(g) Deduce from the previous item that $\hat{\theta}_n$ is consistent.

(h) What is the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$?

Suppose that $n = 100$ and that the observations are such that $\hat{\theta}_n = 3.14$.

(i) Give a numerical approximate two-sided 95% confidence interval for $\theta$.

(j) Consider the testing problem $H_0 : \theta = 3$ against $H_A : \theta \neq 3$ at significance level 5%. Will $H_0$ be rejected?

54. Consider an IID sample $X_1, \ldots, X_n$ from a distribution with mean $\mu$ and finite variance $\sigma^2$, along with another IID sample $Y_1, \ldots, Y_m$ from another distribution with same mean $\mu$, but with variance $2\sigma^2$. The two samples are independent as well. Using the samples separately, one can estimate $\mu$ by the averages $\overline{X}$ and $\overline{Y}$. But we can also mix the estimates and look at $\hat{\mu} = t\overline{X} + s\overline{Y}$, where $t$ and $s$ are some real numbers.

(a) Show that $\hat{\mu}$ is an unbiased estimator of $\mu$ iff $t + s = 1$. *In what follows we assume that* $\hat{\mu}$ *is unbiased!*

(b) Show that the mean squared error of $\hat{\mu}$ is equal to $\frac{t^2\sigma^2}{n} + \frac{2(1-t)^2\sigma^2}{m}$.

(c) The aim is to find the 'best' estimator $\mu$, using an obvious criterion. Show that this best estimator is obtained for $t = \frac{2n}{2n+m}$ and that its mean squared error is equal to $\frac{2\sigma^2}{2n+m}$.

55. Let $X$ be a random variable with a distribution whose probability mass function, depending on an unknown parameter $\theta > 0$, is given by

$$\mathbb{P}_\theta(X = k) = c(\theta)\left(\frac{\theta}{\theta+1}\right)^k,\ k = 0, 1, 2, \ldots$$

Here $c(\theta)$ is an appropriate constant. We consider an IID sample $X_1, \ldots, X_n$ from this distribution.

(a) Show that one must have $c(\theta) = \frac{1}{\theta+1}$. (*Recall the sum of a geometric series!*)

(b) Use the rule $\sum_{k=0}^{\infty} k\alpha^{k-1} = \frac{1}{(1-\alpha)^2}$ for $0 < \alpha < 1$ to show that $\mathbb{E}\, X = \theta$. *N.B. It can also be shown that* $\mathbb{V}\mathrm{ar}\, X = \theta(\theta+1)$, *useful later on.*

(c) Show that the log-likelihood given the sample is $\ell(\theta) = \sum_{i=1}^n X_i \log\theta - (\sum_{i=1}^n X_i + n)\log(\theta+1)$, and show that the sample average $\overline{X}$ is the maximum likelihood estimator of $\theta$. Check that it is indeed a maximum of the likelihood.

(d) Show that the Fisher information $I(\theta)$ in one observation satisfies $I(\theta) = \frac{1}{\theta(\theta+1)}$.

(e) Is the MLE efficient in the Cramér-Rao sense?

56. In this exercise we look at the Weibull distribution. It is a distribution on $(0, \infty)$ with density (for $x > 0$)

$$f_\lambda(x) = \frac{1}{\lambda^k} k x^{k-1} \exp\left(-\frac{x^k}{\lambda^k}\right),$$

where $\lambda > 0$ is the unknown parameter and $k > 0$ a fixed, known, constant. One can show that

$$\mathbb{E}\, X^p = \lambda^p \Gamma\left(\frac{p}{k} + 1\right),$$

if $X$ has the Weibull distribution. Recall that $\Gamma(m) = (m-1)!$, if $m$ is a positive integer.

We have an IID sample $X_1, \ldots, X_n$ from this distribution, want to estimate $\lambda$ and also consider hypothesis testing. It is a fact that $\frac{\sum_{i=1}^n X_i^k}{\lambda^k}$ has the Gamma$(n, 1)$ distribution.

(a) Let $\theta = \lambda^k$. Show that the MLE of $\theta$ is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i^k$.

(b) Compute the MLE $\hat{\lambda}$ of $\lambda$.

(c) Show that $\mathbb{V}\mathrm{ar}\,\hat{\theta} = \frac{\lambda^{2k}}{n}$. Why is $\hat{\theta}$ a consistent estimator of $\theta$?

(d) Deduce from the previous item that $\hat{\lambda}$ is a consistent estimator of $\lambda$.

(e) Compute the log of the likelihood ratio for samples $X_1, \ldots, X_n$ from two Weibull distributions given by parameters $\lambda_0$ and $\lambda_1$. Conclude that for the testing problem $H_0 : \lambda = \lambda_0$ against $H_A : \lambda = \lambda_1$ with $\lambda_1 > \lambda_0$ the likelihood ratio tests rejects $H_0$ for large values of $S := \sum_{i=1}^n X_i^k$. [So, $H_0$ is rejected if $S > s_\alpha$, with $\mathbb{P}_{\lambda_0}(S > s_\alpha) = \alpha$, the significance level.]

(f) Show that $s_\alpha = \lambda_0^k \gamma_\alpha$, where $\gamma_\alpha$ is such that $\mathbb{P}(G > \gamma_\alpha) = \alpha$, if $G$ has a Gamma$(n, 1)$ distribution.

(g) Explain why the likelihood ratio test is uniformly most powerful for the testing problem $H_0 : \lambda = \lambda_0$ against $H_A : \lambda > \lambda_0$, at significance level $\alpha$.

(h) Let $\lambda < \lambda_0$ and assume that the $X_i$ have a Weibull distribution with parameter $\lambda$. Show that $\mathbb{P}_\lambda(S > s_\alpha) \leq \alpha$.

(i) Show that the significance level of the above likelihood ratio test for the testing problem $H_0 : \lambda \leq \lambda_0$ against $H_A : \lambda > \lambda_0$ is again equal to $\alpha$.

57. Consider a regression model in vector form $Y_1 = X\beta + \varepsilon_1$. Here we have that the column *vector* $Y_1$ has $n$ elements and the parameter vector $\beta$ has length $p$. Obviously the design matrix $X$ is of dimensions $n \times p$. The vector $\varepsilon_1$ consists of $n$ independent normal $N(0, \sigma^2)$ random variables.
Next to the above model we also have $Y_2 = 2X\beta + \varepsilon_2$. Here we have that the column vector $Y_2$ again has $n$ elements and the parameter vector $\beta$ and the matrix $X$ are the same as above. The vector $\varepsilon_2$ consists again of independent normal $N(0, \sigma^2)$ random variables, and the vectors $\varepsilon_1$ and $\varepsilon_2$ are also independent.

It is assumed that both vectors $Y_1$ and $Y_2$, as well as the matrix $X$ are observed.

(a) Show that the two above models can be jointly summarized as $\tilde{Y} = \tilde{X}\beta + \varepsilon$ (vectors and matrices of appropriate dimensions), where you express all quantities occurring in this equation in those given above. Show also that $\tilde{X}^\top \tilde{X} = 5X^\top X$.

(b) Show that the least squares estimator $\hat{\beta}$ of $\beta$ is now given by $\hat{\beta} = \frac{1}{5}(X^\top X)^{-1}X^\top(Y_1 + 2Y_2)$. Is it unbiased?

(c) Show that the covariance matrix of $\hat{\beta}$ is $\frac{\sigma^2}{5}(X^\top X)^{-1}$. The ordinary least squares estimator of $\beta$ when only the vector $Y_1$ is observed is $(X^\top X)^{-1}X^\top Y_1$. Is $\hat{\beta}$ as in (b) a better estimator than $(X^\top X)^{-1}X^\top Y_1$?

(d) Let $\hat{\varepsilon} = \tilde{Y} - \tilde{X}\hat{\beta}$, $\hat{\sigma} = \sqrt{\frac{\hat{\varepsilon}^\top\hat{\varepsilon}}{2n-p}}$. Some more notation follows. By $\hat{\beta}_i$ and $\beta_i$ we denote the $i$-th elements of $\hat{\beta}$ and $\beta_i$ respectively, and $(X^\top X)^{-1}_{ii}$ is the $ii$-element of the matrix $(X^\top X)^{-1}$. Put

$$T_i = \frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}\sqrt{(X^\top X)^{-1}_{ii}/5}},$$

and argue (rely on known results) that $T_i$ has a $t$-distribution. With how many degrees of freedom?

(e) Assume $n = 16$, $p = 2$. Suppose one has computed $\hat{\beta}_1 = 1.5$, $\hat{\sigma} = 0.6$ and that $(X^\top X)^{-1}_{11} = 9.8$. Give a 98%-confidence interval for $\beta_1$.
Is the null hypotheses $H_0 : \beta_1 = 1$ to be rejected in favour of the alternative $H_A : \beta \neq 1$ at significance level $\alpha = 0.02$?

58. Consider an IID sample $X_1, \ldots, X_n$ from a distribution with mean $\mu$ and finite variance $\sigma^2$, along with another IID sample $Y_1, \ldots, Y_m$ from another distribution with same mean $\mu$, but with variance $\tau^2$. The two samples are independent as well. The parameter $\mu$ is unknown. Using the samples separately, one can estimate $\mu$ by the averages $\overline{X}$ and $\overline{Y}$. But we can also mix the estimates and look at $\hat{\mu} = \sum_{i=1}^n a_i X_i + \sum_{j=1}^m b_j Y_j$ for certain real numbers $a_i$ and $b_j$.

(a) Show that $\hat{\mu}$ is an unbiased estimator of $\mu$ iff $\sum_{i=1}^n a_i + \sum_{j=1}^m b_j = 1$.

*In what follows we assume that $\hat{\mu}$ is unbiased!*

(b) Show that the mean squared error (MSE) of $\hat{\mu}$ is equal to $\sigma^2 \sum_{i=1}^n a_i^2 + \tau^2 \sum_{j=1}^m b_j^2$.

(c) Find the values of $a_i$ and $b_j$ that makes $\hat{\mu}$ the 'best' estimator in an obvious sense: minimize the MSE under the constraint $\sum_{i=1}^n a_i + \sum_{j=1}^m b_j = 1$ and use 'Lagrange'. Show that this best estimator is obtained for $a_i = \frac{\tau^2}{\tau^2 n + \sigma^2 m}$ and $b_j = \frac{\sigma^2}{\tau^2 n + \sigma^2 m}$.

(d) If $\sigma^2 < \tau^2$ then $a_i > b_j$. Give an intuitive argument why this is not an unreasonable result.

(e) Show that the resulting mean squared error is equal to $\frac{\sigma^2\tau^2}{\tau^2 n+\sigma^2 m}$.

(f) The above estimator hardly deserve this name when $\sigma^2$ and $\tau^2$ are unknown (the $a_i$ and $b_j$ depend on them). But if it would be known that $\sigma^2 = \tau^2$ the problem disappears. Explain why, and show that in this case the mean squared error equals $\frac{\sigma^2}{n+m}$. Why can't this be a surprising result?

59. Consider an IID sample $X_1, \ldots, X_n$ from a distribution with density $f_\theta(x) = \frac{1}{2}\exp(-|x-\theta|)$ ($x \in \mathbb{R}$), where $\theta$ is an unknown parameter. Note that $f_\theta$ is symmetric around $\theta$. Let $\overline{X}_n$ be the sample average of the $X_i$, and $M_n$ the *sample median*. We assume that $n$ is odd, then there are exactly $\frac{1}{2}(n-1)$ observations smaller than $M_n$ and $\frac{1}{2}(n-1)$ observations greater than $M_n$; this explains the name median. It is known that in the present situation $\sqrt{n}(M_n - \theta)$ has an asymptotic standard normal distribution.

(a) What is $\mathbb{E}_\theta X_1$?

(b) Show that $\frac{1}{2}\int_{-\infty}^{\infty}(x-\theta)^2\exp(-|x-\theta|)\,\mathrm{d}x = \int_0^\infty y^2\exp(-y)\,\mathrm{d}y$. Deduce that $\mathbb{V}\mathrm{ar}_\theta(X_1) = 2$.

(c) What is the limit distribution of $\sqrt{n}(\overline{X}_n - \theta)$?

(d) Write $M_n - \theta = \frac{1}{\sqrt{n}}\sqrt{n}(M_n - \theta)$ and deduce that $M_n$ is a consistent estimator of $\theta$.

We consider a testing problem $H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1$, where $\theta_1 > \theta_0$. We consider two test statistics, $T_n^1 = \sqrt{n}(\overline{X}_n - \theta_0)$ and $T_n^2 = \sqrt{n}(M_n - \theta_0)$. Consequently, we have two one-sided tests, both carried out with the same (asymptotic) confidence level $\alpha$. Using $T_n^1$, we reject $H_0$ if $T_n^1 > z_\alpha\sqrt{2}$, and with $T_n^2$ we reject $H_0$ if $T_n^2 > z_\alpha$. [Here $z_\alpha$ is the upper $\alpha$-quantile of the standard normal distribution.]

(e) Show that the test with $T_n^2$ has indeed asymptotic confidence level $\alpha$.

(f) Show that the powers of the two tests in the alternative $\theta_1$ are $1 - \Phi(z_\alpha - (\theta_1 - \theta_0)\sqrt{n})$ and $1 - \Phi(z_\alpha - (\theta_1 - \theta_0)\sqrt{n/2})$.

(g) Which of the two tests is most powerful? Is it also uniformly the more powerful one for the testing problem $H_0 : \theta = \theta_0$ against $H_A : \theta > \theta_0$?

We now consider the testing problem $H_0 : \theta = \theta_0$ against $H_A : \theta \neq \theta_0$, with the same test statistics and the same common significance level $\alpha$ as above.

(h) Give two $(1-\alpha)$-confidence intervals for $\theta_0$, based on $T_n^1$ and $T_n^2$ respectively.

(i) Suppose it happened that both $\overline{X}_n$ and $M_n$ are equal to 0.6 and that $\alpha$ and $n$ are such that $z_{\alpha/2}/\sqrt{n} = 0.1$. Which of the test rejects $H_0 : \theta_0 = 0.48$?

60. Consider the standard linear regression model, in unusual vector notation, $Y = Z\gamma + \varepsilon$. Here, as usual, $\varepsilon$ is a $n$-dimensional vector whose components are IID with a common $N(0, \sigma^2)$ distribution. Recall that the covariance matrix of $\varepsilon$ is $\sigma^2 I_n$, with $I_n$ the $n$-dimensional identity matrix. The vector $\gamma$ is the $q$-dimensional parameter, $q \leq n$ and $Z$, of size $n \times q$, is the matrix of regressors. The idea is to perform the usual least squares estimation procedure to estimate $\gamma$, but the problem is that $Z$ doesn't have full rank, its rank is $p < q$. Consequently, the matrix $V := Z^\top Z$ is not invertible. One says that the model is overparametrized and as a result, as we shall see, there is no good estimator of $\gamma$.
Linear algebra tells us that we can write $Z = XA$, where $X \in \mathbb{R}^{n\times p}$ and $A \in \mathbb{R}^{p\times q}$, where both $X$ and $A$ have rank $p$. Then $X^\top X$ is invertible and there exists a matrix $B \in \mathbb{R}^{q\times p}$ such that $AB = I_p$, the $p$-dimensional identity matrix. Note that $BA$ is not equal to $I_q$ and that $X, A, B$ all depend on the elements of $Z$, so there are in principle computable. Furthermore we define $\beta := A\gamma \in \mathbb{R}^p$ (then $Z\gamma = X\beta$) and $V^+ := B(X^\top X)^{-1}B^\top$.
We estimate $\gamma$, in analogy to the usual case, by $\hat{\gamma} := V^+ Z^\top Y$.

(a) Show that $V^+V = BA$, $VV^+ = (BA)^\top$ and $V^+VV^+ = V^+$ [useful for further computations].

(b) Show that $\hat{\gamma} = BA\gamma + V^+Z^\top\varepsilon$. Is $\hat{\gamma}$ an unbiased estimator of $\gamma$?

(c) Show that the covariance matrix of $\hat{\gamma}$ is equal to $\sigma^2 V^+$.

(d) In this context, an obvious estimator of $\beta$ is $\hat{\beta} = A\hat{\gamma}$. Show that $\hat{\beta}$ is an unbiased estimator of $\beta$ and that its covariance matrix is $\sigma^2(X^\top X)^{-1}$, as in the usual case.

(e) Let $\hat{Y}$ be the prediction of $Y$ using $\hat{\gamma}$, $\hat{Y} = Z\hat{\gamma}$. Show that $ZV^+Z^\top = X(X^\top X)^{-1}X^\top$ and that $\hat{Y} = X\beta + X(X^\top X)^{-1}X^\top\varepsilon$, as usual.

(f) The residual vector is $\hat{\varepsilon} = Y - \hat{Y}$. Show that $\mathbb{E}\hat{Y} = Z\gamma$ and that $\hat{\varepsilon} = (I - X(X^\top X)^{-1}X^\top)\varepsilon$. [We can thus use $S^2 = ||\hat{\varepsilon}||^2/(n-p)$ as the usual unbiased estimator of $\sigma^2$, useful below.]

(g) Although there is no sensible way to give a confidence interval for elements of $\gamma$, we can still construct a confidence interval for elements of the vector $\mathbb{E}Y = Z\gamma$. We concentrate on the first element of $\mathbb{E}Y$, call it $\mu_0$ and note that $\mu_0 = e^\top Z\gamma$, where $e^\top = (1, 0, \ldots, 0)$. Show that the random variable $\hat{\mu}_0 := e^\top\hat{Y}$ has a normal distribution with mean $\mu_0$ and variance $\sigma_0^2 := \sigma^2 e^\top ZV^+Z^\top e$.

(h) Show that $\hat{\mu}_0 \pm t_{n-p,\alpha/2} S \sqrt{e^\top Z V^+ Z^\top e}$ are the bounds of a $(1-\alpha)$-confidence interval for $\mu_0$. [Here $t_{n-p,\alpha/2}$ is the usual $\alpha/2$ upper quantile of the $t_{n-p}$-distribution.]

61. Here we consider IID random variables $X, X_1, \ldots, X_n$ that have a common Poisson($\lambda$) distribution. We will estimate $\theta = \lambda^2$ based on the sample $X_1, \ldots, X_n$.

(a) Show that $\mathbb{E}\,(X(X-1)\cdots(X-k+1)) = \lambda^k$ for an integer $k \geq 1$. *Hint: write the expectation as a sum of values $n(n-1)\cdots(n-k+1)$ times the probabilities $\mathbb{P}(X=n) = \lambda^n e^{-\lambda}/n!$ and see how this impacts on $n!$*

(b) Show that $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i(X_i - 1)$ is an unbiased estimator of $\theta$.

(c) Let $i \neq j$. Show that $\mathbb{E}\,(X_i X_j (X_i - 1)(X_j - 1)) = \lambda^4$. What is the covariance $\mathbb{C}\mathrm{ov}(X_i(X_i-1), X_j(X_j-1))$?

(d) Verify that $X^2(X-1)^2 = X(X-1)(X-2)(X-3) + 4X(X-1)(X-2) + 2X(X-1)$, and show that $\mathbb{E}\,(X^2(X-1)^2) = \lambda^4 + 4\lambda^3 + 2\lambda^2$.

(e) Show that $\mathbb{V}\mathrm{ar}\,(X(X-1)) = 4\lambda^3 + 2\lambda^2$ and that $\mathbb{V}\mathrm{ar}\,(\widehat{\theta}_n) = \frac{4\theta^{3/2}+2\theta}{n}$.

(f) Is $\widehat{\theta}_n$ a consistent estimator of $\theta$?

(g) Show that the Fisher information $I(\theta) = \frac{1}{4\theta^{3/2}}$ (to do this, first you rewrite the probability $\mathbb{P}(X=n)$ in terms of $\theta$). Give an explicit formula for an estimator $\widetilde{\theta}_n$ of $\theta$ that asymptotically has variance $\frac{4\theta^{3/2}}{n}$. [*Recall that the MLE of $\lambda$ is the sample mean.*]

(h) Which of the estimators $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ would you prefer, and why?

62. Consider the standard multivariate regression model $Y = X\beta + \varepsilon$ ($Y$ is $n$-dimensional, $X \in \mathbb{R}^{n\times p}$, $\beta = \left(\beta_0, \beta_1, \ldots, \beta_{p-1}\right)^\top \in \mathbb{R}^p$, $\mathbb{C}\mathrm{ov}(\varepsilon) = \sigma^2 I$ and all other common assumptions). Let $\eta = A\beta$ where $A$ is a known invertible matrix. Then we also have the alternative model $Y = U\eta + \varepsilon$ for $U = XA^{-1}$.

(a) Compute for both models the least squares estimators $\hat{\beta}$ and $\hat{\eta}$ (in terms of $X$, $U$ and $Y$) and show that that $\hat{\eta} = A\hat{\beta}$.

(b) Consider the usual estimators of $\sigma^2$ for each of the two models. Show that they are the same.

We are interested in testing the null hypothesis $\beta_0 = \beta_1$ against the alternative $\beta_0 \neq \beta_1$ at a significance level $\alpha$. We let $A$ be the $p \times p$ matrix which is the identity matrix except for its first row, which is $\begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \end{pmatrix}$.

(c) Give a test statistic for the above testing problem, call it $T$, and give its distribution under the null hypothesis.

Suppose $n = 12$, $p = 2$. The least squares estimators are $\hat{\beta} = \begin{pmatrix} 2.97 \\ 3.49 \end{pmatrix}$. You may also want to use $(X^\top X)^{-1} = \begin{pmatrix} 0.0098 & -0.0070 \\ -0.0070 & 0.0067 \end{pmatrix}$ and $\hat{\sigma} = 1.592$.

(d) Compute $\begin{pmatrix} 1 & -1 \end{pmatrix} (X^\top X)^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Give the rejection region of the test statistic $T$ and perform the test with the above data and $\alpha = 0.05$.

(e) Give a theoretical and numerical 95% confidence region for the parameter $\beta_0 - \beta_1$. Is this result in agreement with the answer to previous question?

63. In this question you have to compute most answers by performing calculations with (double) integrals. Let $(X, Y)$ be a random vector whose density (depending on a parameter $\theta > 0$) is given by

$$f_\theta(x, y) = c(\theta) \exp(-\frac{|x|}{\theta} - \frac{|y|}{\theta}) \mathbf{1}_{xy>0},$$

where $c(\theta)$ is the positive normalization constant. Note that $f_\theta(x, y) = 0$ for all $x, y$ with $xy \leq 0$ and so $\mathbb{P}(XY > 0) = 1$.

(a) Show that $c(\theta) = \frac{1}{2\theta^2}$. [*You can write the double integral as a sum of two integrals, one for* $x, y > 0$ *and one for* $x, y < 0$.]

(b) Show that $X$ has marginale density $f_\theta(x) = \frac{1}{2\theta} \exp(-\frac{|x|}{\theta})$. You have to do this for one of the two cases $x \geq 0$ and $x < 0$ only.

(c) Argue without computation that $\mathbb{E}\, X = 0$. Give also the marginal density of $Y$.

(d) Show that $\mathbb{E}\,|X| = \theta$.

(e) Show that $\mathbb{E}\, X^2 = 2\theta^2$. What are $\mathbb{V}\mathrm{ar}\, X$ and $\mathbb{V}\mathrm{ar}\, |X|$?

(f) Show that $\mathbb{E}\,(XY)$ can be computed as $\mathbb{E}\,(XY) = \frac{1}{\theta^2}(\int_0^\infty x \exp(-\frac{x}{\theta})\, \mathrm{d}x)^2$ with $\mathbb{E}\,(XY) = \theta^2$ as a result.

(g) Are $X$ and $Y$ independent random variables?

(h) Show that $\mathbb{E}\,|XY| = \theta^2$. What is $\mathbb{C}\mathrm{ov}(|X|, |Y|)$?

64. We consider random variables $X$ having a Weibull distribution, a distribution on $(0, \infty)$ with density (for $x > 0$)

$$f_\theta(x) = c(\theta, k) x^{k-1} \exp\left(-\frac{x^k}{\theta}\right),$$

where $\theta > 0$ is an unknown parameter, $k > 0$ a fixed, known, positive constant and $c(\theta, k)$ a normalization constant.

We also have an IID sample $X_1, \dots, X_n$ from this distribution, want to estimate $\theta$ and consider hypothesis testing. For future use we mention that $\frac{1}{\theta}\sum_{i=1}^n X_i^k$ has the Gamma$(n, 1)$ distribution.

(a) Show that $\mathbb{E}\, X^p = \frac{c(\theta,k)}{k}\theta^{\frac{p}{k}+1}\Gamma\left(\frac{p}{k}+1\right)$ for $p > 0$.

(b) Show by a direct computation that for $f_\theta$ to be a probability density function, one needs $c(\theta, k) = \frac{k}{\theta}$.

(c) Show that the maximum likelihood estimator of $\theta$ is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i^k$.

(d) Show that $\mathbb{V}\mathrm{ar}\, \hat{\theta} = \frac{\theta^2}{n}$. Why is $\hat{\theta}$ a consistent estimator of $\theta$?

(e) Compute the log of the likelihood ratio for samples $X_1, \dots, X_n$ from two Weibull distributions given by parameters $\theta_0$ and $\theta_1$. Conclude that for the testing problem $H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1$ with $\theta_1 > \theta_0$ the likelihood ratio tests rejects $H_0$ for large values of $S := \sum_{i=1}^n X_i^k$. [So, $H_0$ is rejected if $S > s_\alpha$, with $\mathbb{P}_{\theta_0}(S > s_\alpha) = \alpha$, the significance level.]

(f) Show that $s_\alpha = \theta_0\gamma_{n,\alpha}$, where $\gamma_{n,\alpha}$ is such that $\mathbb{P}(G_n > \gamma_{n,\alpha}) = \alpha$, if $G_n$ has a Gamma$(n, 1)$ distribution.

(g) Explain why the likelihood ratio test is uniformly most powerful for the testing problem $H_0 : \theta = \theta_0$ against $H_A : \theta > \theta_0$, at significance level $\alpha$.

(h) Let $\theta < \theta_0$ and assume that the $X_i$ have a Weibull distribution with parameter $\theta$. Show that $\mathbb{P}_\theta(S > s_\alpha) < \alpha$, and conclude that $\sup_{\theta\leq\theta_0} \mathbb{P}_\theta(S > s_\alpha) = \alpha$. [This means that also the significance level of the above likelihood ratio test for the testing problem $H_0 : \theta \leq \theta_0$ (now a composite hypothesis) against $H_A : \theta > \theta_0$ is equal to $\alpha$.]

(i) Is this likelihood ratio test also uniformly most powerful for the testing problem $H_0 : \theta \leq \theta_0$ against $H_A : \theta > \theta_0$, at significance level $\alpha$.

65. Here we consider the standard multivariate regression model, but in two versions. The first one is given by the familiar equation

$$Y = X\beta + \varepsilon, \tag{3}$$

with all usual assumptions are satisfied, like $\beta \in \mathbb{R}^p$, $\beta^\top = (\beta_0, \dots, \beta_{p-1})$, $X \in \mathbb{R}^{n\times p}$, $\varepsilon$ a vector of $n$ IID random variables with a $N(0, \sigma^2)$ distribution. The second model is given by

$$Y = Z\beta^e + \varepsilon, \tag{4}$$

where $\beta^e \in \mathbb{R}^{p+1}$ is the extended vector $\begin{pmatrix} \beta \\ \beta_p \end{pmatrix}$, in which $\beta_p \in \mathbb{R}$ is an additional parameter, and $Z = \begin{pmatrix} X & X_p \end{pmatrix} \in \mathbb{R}^{n\times(p+1)}$, where $X_p \in \mathbb{R}^n$ is an additional vector, $X_p^\top = (x_{1,p}, \ldots, x_{n,p})$. The second model can then also be represented as $Y = X\beta + X_p\beta_p + \varepsilon$. For both models we consider the corresponding least squares estimators, $\hat\beta = (X^\top X)^{-1}X^\top Y$ and $\hat\beta^e = (Z^\top Z)^{-1}Z^\top Y$. All inverses are assumed to exist, which happens if $Z$ has full rank. We investigate some consequence of using two different models.

(a) If $\beta_p$ is known in the second model (6), there is no need to estimate it. Suppose $\beta$ is still unknown and to be estimated. A least squares type of estimator of $\beta$ in this case, different from the one that can be obtained from $\hat\beta^e$ as mentioned above, is $\tilde\beta := (X^\top X)^{-1}X^\top(Y - X_p\beta_p)$. Show that this estimator is unbiased and give its covariance matrix. Would this be a sensible estimator of $\beta$ if $\beta_p$ is unknown?

Obviously both models can't be true at the same time (unless $\beta_p$ is known to be zero). We are therefore interested in properties of the estimators under misspecification. We consider two cases, the first one is to assume *the second model to be true* so all computations have to be done according to Equation (6), but we use the ordinary least squares estimator $\hat\beta$ as an estimator of $\beta$.

(b) Assume model (6) and estimate $\beta$ by $\hat\beta$. Show that the bias $\mathbb{E}\,\hat\beta - \beta$ is equal to $(X^\top X)^{-1}X^\top X_p\beta_p$ and that the covariance matrix of $\hat\beta$ equals $\sigma^2(X^\top X)^{-1}$.

(c) Show that the mean squared error of a $\hat\beta_i$ for $i = 0, \ldots, p-1$ can be computed as $\sigma^2 e_i^\top (X^\top X)^{-1}e_i + \big(e_i^\top(X^\top X)^{-1}X^\top X_p\big)^2\,\beta_p^2$, where $e_i \in \mathbb{R}^p$ is an appropriate vector having one entry equal to one, the others zero.

To be on the safe side when it is not clear which of the two models is true, one can always assume the extended model (6) and estimate $\beta$ by taking the first $p$ elements of the vector $\hat\beta^e$. Call the resulting estimator $\tilde\beta$ (although the notation is the same, it is not the estimator of item (a)). One has $\tilde\beta = \begin{pmatrix} I_p & 0 \end{pmatrix}\hat\beta^e$, where the zero in the matrix is to be understood as a zero vector of length $p$. Even when the first model, Equation (5), is the true one we use $\tilde\beta$ as an estimator of $\beta$ (instead of $\hat\beta$).

(d) Show that $\tilde\beta$ is an unbiased estimator of $\beta$ and that the covariance matrix of $\tilde\beta$ is the upper left $p \times p$ block of $\sigma^2(Z^\top Z)^{-1}$.

(e) The upper left block in the previous item is difficult to compute, but it turns out that

$$\mathbb{C}\mathrm{ov}(\tilde{\beta}) = \sigma^2 (X^\top X - \frac{1}{X_p^\top X_p} X^\top X_p X_p^\top X)^{-1}.$$

It is then possible to show (which you don't have to do, unless you really want) that for the diagonal elements of the matrices one has $\mathbb{C}\mathrm{ov}(\tilde{\beta})_{ii} \geq \mathbb{C}\mathrm{ov}(\hat{\beta})_{ii}$. Is it wise to be on the safe side if you know that Equation (5) is the true model?

(f) The two models coincide if $\beta_p = 0$. It is therefore useful to consider the testing problem $H_0 : \beta_p = 0$ against the alternative $H_A : \beta_p \neq 0$ at a significance level $\alpha$. Give a test statistic for this testing problem, its distribution and use this to give a $(1-\alpha)$-confidence interval for $\beta_p$.

66. We have observations in the form of $n$ IID stochastic vectors $(X_1, Y_1), \ldots, (X_n, Y_n)$, all having the same (joint) density (depending on a parameter $\theta > 0$) as the stochastic vector $(X, Y)$. This density is given by

$$f_\theta(x, y) = \frac{1}{2\theta^2} \exp\left(-\frac{|x|}{\theta} - \frac{|y|}{\theta}\right) \mathbf{1}_{xy>0}.$$

The marginal density of $X$ (and by symmetry also of $Y$ !) turns out to be

$$p_\theta(x) = \frac{1}{2\theta} \exp(-\frac{|x|}{\theta}).$$

Furthermore it is known that the absolute values $|X|$ and $|Y|$ are **independent**.

We will estimate $\theta$ by the maximum likelihood method, and write $\hat{\theta}_n$ for the maximum likelihood estimator based on the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$.

(a) Are $X$ and $Y$ independent?

(b) Show that $|X|$ has density $g_\theta(x) = \frac{1}{\theta} \exp(-\frac{x}{\theta})$ for $x \geq 0$, which should look familiar. [*Hint: compute first as an integral the probability* $\mathbb{P}_\theta(|X| \leq x) = \mathbb{P}_\theta(-x \leq X \leq x)$ *for* $x \geq 0$. *You may want to exploit a certain symmetry too.*]

(c) Show that $\mathbb{E}_\theta(|X|) = \theta$ and $\mathbb{V}\mathrm{ar}_\theta(|X|) = \theta^2$, preferably without cumbersome computations.

(d) Show that the likelihood is given by

$$L(\theta) = 2^{-n} \theta^{-2n} \exp\left(-\frac{\sum_{i=1}^n |X_i| + \sum_{i=1}^n |Y_i|}{\theta}\right) \prod_{i=1}^n \mathbf{1}_{X_i Y_i > 0}.$$

(e) Show that $\hat{\theta}_n = \frac{1}{2n}(\sum_{i=1}^n |X_i| + \sum_{i=1}^n |Y_i|)$. Is $\hat{\theta}_n$ an unbiased estimator of $\theta$?

(f) Argue by the Central limit theorem that $\sqrt{2n}(\frac{\hat{\theta}_n}{\theta} - 1)$ has an approximate standard normal distribution for large $n$. [*Hint: Note that $\hat{\theta}_n$ is the average of the $\frac{1}{2}(|X_i| + |Y_i|)$.* ]

(g) Show that the Fisher information in one observation $(X, Y)$ about $\theta$ is $\frac{2}{\theta^2}$.

(h) Use the Fisher information and a result for maximum likelihood estimators to establish that $\sqrt{2n}(\frac{\hat{\theta}_n}{\theta} - 1)$ has an approximate standard normal distribution for large $n$.

(i) Consider the testing problem $H_0 : \theta = \theta_0$ (null hypothesis) against the alternative $H_A : \theta = \theta_1$ at the significance level $\alpha$, where $\theta_1 > \theta_0$. Show that the likelihood ratio test rejects the null hypothesis for large values of $\hat{\theta}_n$, i.e. if $\hat{\theta}_n > c$ for some $c > 0$.

(j) Show that the critical value $c$ in the preceding question can be approximated (for large $n$) by $\theta_0(1+\frac{z_\alpha}{\sqrt{2n}})$, where $z_\alpha$ is such that $\mathbb{P}(Z > z_\alpha) = \alpha$ when $Z$ has a standard normal distribution.

(k) Is the likelihood ratio test of item (i) uniformly most powerful for the testing problem $H_0 : \theta = \theta_0$ against the alternative $H_A : \theta < \theta_0$ (carefully note the inequality here!) at the same significance level $\alpha$?

(l) Let $z = z_{\alpha/2}$ be such that $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$ when $Z$ has a standard normal distribution. Show that $(\frac{\hat{\theta}_n}{1+\frac{z}{\sqrt{2n}}}, \frac{\hat{\theta}_n}{1-\frac{z}{\sqrt{2n}}})$ is an approximate $(1-\alpha)$-confidence interval for $\theta$.

(m) Suppose that in practical situation with $n = 50$ one finds $\hat{\theta}_n = 3.25$ and that $\alpha = 0.10$. Should the null hypothesis $\theta = \pi$ (the famous number $\pi$) be rejected in favour of the alternative $\theta \neq \pi$ when using a test based on $\hat{\theta}_n$?

67. Let $X_1, \ldots, X_n$ be IID having a $U[\theta, 2\theta]$ distribution (uniform on the interval $[\theta, 2\theta]$) with $\theta > 0$. Note that for such random variables it holds that $X_i \geq \theta$ and $X_i \leq 2\theta$, more precisely $\mathbb{P}_\theta(X_i \geq \theta,\, X_i \leq 2\theta) = 1$, where we use the notation $\mathbb{P}_\theta$ to emphasize that we compute probabilities under the parameter $\theta$. Let $M_n = \max\{X_1, \ldots, X_n\}$, $m_n = \min\{X_1, \ldots, X_n\}$ and $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

(a) Show that $\mathbb{P}_\theta(\frac{1}{2}M_n \leq m_n) = 1$ for all $\theta > 0$.

(b) Show that the likelihood can be written as

$$L(\theta) := \frac{1}{\theta^n}\mathbf{1}_{\theta \leq m_n}\mathbf{1}_{\theta \geq \frac{1}{2}M_n}.$$

(c) Sketch the graph of $L$ as a function of $\theta$ (taking $n = 1$ already gives a good impression) and argue that the maximum likelihood estimator (MLE) of $\theta$ is $\hat{\theta}_n := \frac{1}{2}M_n$.

(d) For consistency of the MLE one has to show two things for all small $\varepsilon > 0$: $\mathbb{P}_\theta(\hat{\theta}_n < \theta - \varepsilon) \to 0$ and $\mathbb{P}_\theta(\hat{\theta}_n > \theta + \varepsilon) \to 0$ for $n \to \infty$. Show one of these.

(e) Give a consistent estimator of $\theta$ based on $m_n$; motivation is not required.

(f) Another estimator of $\theta$ is $c\,\overline{X}_n$ for some $c$. How to choose $c$ to get an unbiased estimator? Give a brief argument (no extensive computations, refer to a theorem) for consistency of this estimator.

68. Here we consider the standard multivariate regression analysis, but in two versions. In the first situation for the first analysis one has the familiar equation

$$Y = X\beta + \varepsilon, \tag{5}$$

with all usual assumptions are satisfied, like $\beta \in \mathbb{R}^p$, $\beta^\top = (\beta_0, \ldots, \beta_{p-1})$, $X \in \mathbb{R}^{n\times p}$, $\varepsilon$ a vector of $n$ IID random variables with a $N(0, \sigma^2)$ distribution. The second analysis, carried out **independently** of the first one, uses the same model setup but with different data (even their number $n'$ may be different), and is given by

$$Y' = X'\beta + \varepsilon', \tag{6}$$

again with all usual assumptions are satisfied, like $\beta \in \mathbb{R}^p$, $\beta^\top = (\beta_0, \ldots, \beta_{p-1})$, $X' \in \mathbb{R}^{n'\times p}$, $\varepsilon'$ a vector of $n'$ IID random variables with a $N(0, \sigma^2)$ distribution. Note that in both models one has the **same** $\beta$ and the **same** $\sigma^2$ and that the prime in e.g. $X'$ is only used to distinguishing notation, nothing do with derivatives or transposition.
The random variables used in the different versions are assumed to be independent, so $Y$ and $Y'$ are **independent** random vectors, as wel as $\varepsilon$ and $\varepsilon'$.

For both models we consider the corresponding least squares estimators of the common parameter $\beta$, $\hat{\beta} = (X^\top X)^{-1}X^\top Y$ and $\hat{\beta}' = (X'^\top X')^{-1}X'^\top Y'$. Both are unbiased estimators of $\beta$; their covariance matrices are denoted $C$ and $C'$ respectively and are known from the theory, e.g. $C = \sigma^2(X^\top X)^{-1}$. All inverses are assumed to exist. We investigate some consequence of using two different models and how to combine them.

(a) A combined estimator is $A\hat{\beta} + B\hat{\beta}'$, where $A$ and $B$ are $p \times p$ matrices. Show that the combined estimator is unbiased for $B = I - A$. This choice will be made from now on, and we write $\hat{\beta}_A$ for the resulting estimator. Show also that the covariance matrix of $\hat{\beta}_A$ is $C_A := ACA^\top + (I - A)C'(I - A)^\top$.

(b) We make a special choice for $A$, we take $A^* = C'(C + C')^{-1}$. With this choice one has for any $A$ the identity $C_A = C_{A^*} + (A - A^*)(C + C')(A - A^*)^\top$ (this is given information that you don't have to prove). Suppose that one wants to estimate a linear combination $u^\top \beta$, with some know vector $u$. Show that $\hat{u}_A = u^\top \hat{\beta}_A$ is an unbiased of $u^\top \beta$ and that its variance is $u^\top C_A u$. According to what criterion would one call $A^*$ an optimal choice of the estimator $u_A$?

(c) Suppose that in the second experiment the same matrix $X$ is used as in the first one (one can think of an independent repetition of the experiment, but with the same values of the predictor variables), so $X = X'$ (and $n = n'$). Show that $A^* = \frac{1}{2}I$. Give also an explicit expression for $C_{A^*}$ in terms of $X$ and $\sigma^2$ this case.

A supervisor who has access to both experiments and their data sets draws up a simultaneous model using $X' = X$,

$$\begin{pmatrix} Y \\ Y' \end{pmatrix} = \begin{pmatrix} X \\ X \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ \varepsilon' \end{pmatrix}.$$

(d) Show that it follows from general least squares theory *applied to this model* that the least squares estimator of $\beta$ is $\hat{\beta}^s := \frac{1}{2}(X^\top X)^{-1} X^\top (Y + Y')$.

(e) Compute the covariance matrix of $Y + Y'$ and the covariance matrix $\mathbb{C}\mathrm{ov}(\hat{\beta}^s)$ of $\hat{\beta}^s$ and show that the latter one coincides with $C_{A^*}$ above.

(f) The supervisor wants to test the null hypothesis $H_0$ that the last element $\beta_{p-1}$ of the vector $\beta$ is zero against the alternative that it is unequal to zero, using the data from both experiments and the simultaneous model above. Write down a test statistic and specify it's distribution under $H_0$.

(g) In a concrete situation one has $n = 70$, $p = 20$, computes $\hat{\beta}_{p-1} = -0.72$, the bottom right element of $(X^\top X)^{-1}$ equals $1.28$ and $\sigma^2$ is estimated by $0.36$. Will the null hypothesis be rejected at the level $\alpha = 0.05$?

69. A positive random variable $X$ is said to have an *inverse gamma distribution*, denoted $\mathrm{IG}(\alpha, \beta)$, if it has a density

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\frac{1}{x})^{\alpha+1} \exp\left(-\frac{\beta}{x}\right),$$

for $x > 0$ and where $\alpha, \beta$ are positive parameters.

(a) Show that $f(x; \alpha, \beta)$ is indeed a density for every $\alpha, \beta > 0$ by computing its integral. [The substitution $y = \beta/x$ is helpful.]

(b) Show that $\mathbb{E}\, X = \frac{\beta}{\alpha-1}$ for $\alpha > 1$. [You may want to use that $f(x; \alpha', \beta)$ is a density for $\alpha' = \alpha - 1$, and the property $\Gamma(\alpha' + 1) = \alpha' \Gamma(\alpha')$.]

(c) If $X$ has an $\mathrm{IG}(\alpha, \beta)$ distribution then $X/\beta$ has an $\mathrm{IG}(\alpha, 1)$ distribution. Show this by deriving the density of $X/\beta$ from that of $X$.

In the remainder of this exercise $\alpha$ is supposed to be a known constant and $\beta$ an unknown parameter, which is to be estimated from a sample of IID random variables $X_1, \ldots, X_n$ from the $\mathrm{IG}(\alpha, \beta)$ distribution. It is given that $\mathbb{E}\, X_1^k = \beta^k \frac{\Gamma(\alpha-k)}{\Gamma(\alpha)}$ for every $k < \alpha$.

(d) Show that, for $\alpha > 1$, the moment estimator (using the first moment only) of $\beta$ is $\hat{\beta}^1 := (\alpha - 1)\overline{X}_n$, where $\overline{X}_n$ is the sample average. Is $\hat{\beta}^1$ an unbiased estimator of $\beta$?

(e) Show that $\mathbb{V}\mathrm{ar}\, \hat{\beta}^1 := \frac{1}{n} \frac{\beta^2}{\alpha-2}$ for $\alpha > 2$. Is $\hat{\beta}^1$ consistent?

(f) Show that $\mathbb{E}\, \frac{1}{n} \sum_{i=1}^n \frac{1}{X_i} = \frac{\alpha}{\beta}$.

(g) Show that the likelihood of the sample is

$$\frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \prod_{i=1}^n X_i^{-(\alpha+1)} \exp\left(-\beta \sum_{i=1}^n \frac{1}{X_i}\right).$$

and that the maximum likelihood estimator of $\beta$ is

$$\hat{\beta}_n = \frac{\alpha}{\frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}}.$$

(h) Apply the law of large numbers to show consistency of the MLE.

(i) Show that $\mathbb{E}\, \hat{\beta}_n = \frac{n\beta}{n\alpha-1}$.

(j) Show that the Fisher information about $\beta$ (for $n = 1$) equals $\frac{\alpha}{\beta^2}$.

In the remaining part of this exercise we assume that **the parameter $\alpha$ of the $\mathrm{IG}(\alpha, \beta)$ equals 1**, and the symbol $\alpha$ will be reserved to denote significance level.

(k) Use the Likelihood Ratio (LR) test to test the null hypothesis $H_0$ : $\beta = \beta_0$ against the alternative $H_A : \beta = \beta_0'$, where $\beta_0' > \beta_0$, at some significance level $\alpha$. Let $T := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{X_i}$. Show that the LR test rejects the null hypothesis for small values of $T$, say $T \leq c_l$ for some $c_l$, where $c_l$ depends on $\alpha$ and $\beta_0$. We keep on calling the latter test LR test in everything that follows.

(l) Is the LR test also UMP for the testing problem $H_0 : \beta = \beta_0$ against the alternative $H_A : \beta > \beta_0$?

(m) Argue that for all $\beta$ and $X_i$ having an $\mathrm{IG}(1, \beta)$ distribution it holds that $\beta T$ has a distribution that doesn't depend on $\beta$, i.e. $P_\beta(\beta T \leq y)$ doesn't depend on $\beta$ for every $y > 0$, where the notation $P_\beta$ is used to emphasize that the $X_i$ (and hence $T$) have a distribution depending on $\beta$. [Look at question (c).]

(n) Show that, in similar notation, $\mathbb{P}_\beta(T \leq c_l) \leq \alpha$ for all $\beta \leq \beta_0$. [*Show first that* $\mathbb{P}_\beta(T \leq c_l) = \mathbb{P}_{\beta_0}(T\beta_0 \leq \beta c_l)$.] What is the significance level of the LR test for the testing problem $H_0 : \beta \leq \beta_0$ against the alternative $H_A : \beta > \beta_0$?

The significance level of the LR test for a testing problem with $H_0$ : $\beta \leq \beta_0$ is $\sup_{\beta \leq \beta_0} \mathbb{P}_\beta(T \leq c_l)$. Compute this significance level.

70. Let $X$ be a nonnegative random variable with a density $f(x) = \frac{1}{\theta}(x+1)^{-1/\theta - 1} = \lambda(x+1)^{-\lambda-1}$, where $x \geq 0$ and $\theta, \lambda > 0$ are parameters related by $\lambda = \frac{1}{\theta}$.

(a) Show that $\mathbb{E}\,(X+1)^a = \frac{\lambda}{\lambda - a}$ when $a < \lambda$.

(b) Show that $\mathbb{E}\,\log(X+1) = \frac{1}{\lambda}$ for any value of $\lambda$. *Hint:* use the substitution $u = \log(x+1)$ when you compute a relevant integral.

Let $X_1, \ldots, X_n$ be IID with common density $f$, where $\lambda$ and $\theta$ are the unknown parameters to be estimated.

(c) Compute the moment estimator of $\lambda$ based on the sample average. What is the corresponding moment estimator of $\theta$?

(d) Show that the maximum likelihood estimator of $\lambda$ is $\frac{n}{\sum_{i=1}^{n} \log(X_i+1)}$. What is the maximum likelihood estimator of $\theta$?

(e) Which of the two maximum likelihood estimators is unbiased?

(f) Show that the Fisher information about $\lambda$ in one observation is equal to $\frac{1}{\lambda^2}$. [A second derivative might be helpful here.]

(g) Deduce from the previous item that $\mathbb{V}\mathrm{ar}\,(\log(1+X_1)) = \frac{1}{\lambda^2} = \theta^2$.

(h) Compute also the Fisher information about $\theta$ in one observation.

(i) Use the central limit theorem for the mean to compute the bounds of a $(1-\alpha)$-confidence interval for $\theta$.

(j) Use the central limit theorem for the mean to establish that the limit distribution of $\sqrt{n}(\frac{\hat{\theta}_n}{\theta} - 1)$ is standard normal.

(k) Use the previous result to obtain a $(1-\alpha)$-confidence interval for $\theta$.

(l) Based on the asymptotic behavior of the maximum likelihood estimators, it is possible to obtain (approximate) $(1-\alpha)$-confidence intervals. Give such an interval for $\theta$ or for $\lambda$ (one interval suffices).

(m) Use one of the maximum likelihood estimators to test the null hypothesis $H_0 : \lambda = 1$ (or $H_0 : \theta = 1$) against the alternative $H_A : \lambda \neq 1$ (or $H_0 : \theta \neq 1$) at significance level $\alpha$. Describe the critical region for the chosen test statistic.

(n) Suppose $n = 256$ and the MLE $\hat{\theta}_n = 1.25$ and that the significance level is 0.10. Compute for $\theta$ or for $\lambda$ a confidence interval. Should the null hypothesis above be rejected?

71. In this exercise we will consider a variation on the least squares approach to estimation for linear multivariate regression models. The model we consider can be concisely represented by

$$Y = X\beta + \varepsilon,$$

where $Y$ is random $n$-dimensional vector, $X$ an $(n \times p)$-dimensional deterministic (design) matrix and $\varepsilon$ a random $n$-dimensional vector that has a multivariate normal distribution with mean zero (as a vector) and a covariance matrix $\Sigma$ that is invertible. The *modified* least squares estimator $\hat{\beta}$ of $\beta$ that we will consider is the minimizer of $\mathrm{LS}(\beta) := (Y - X\beta)^\top P(Y - X\beta)$, where $P$ is a given symmetric strictly positive definite $n \times n$ matrix. In particular its inverse $P^{-1}$ exists. It turns out that $\hat{\beta} = (X^\top PX)^{-1}X^\top PY$, where it is assumed that $X$ has full rank equal to $p$. Note that $Y - X\hat{\beta} = QY$, where $Q = I - X(X^\top PX)^{-1}X^\top P$.

(a) What is the distribution of $Y$?

(b) What are $X^\top PQ$, $QX$? Show that $Q^\top PQ = P - PX(X^\top PX)^{-1}X^\top P$.

(c) Show that $\mathrm{LS}(\beta) = (\beta - \hat{\beta})^\top X^\top PX(\beta - \hat{\beta}) + \mathrm{LS}(\hat{\beta})$ (this requires some nasty matrix computations). Deduce that $\hat{\beta}$ is the minimizer of $\mathrm{LS}(\beta)$.

(d) Show that $\hat{\beta}$ is an unbiased estimator of the vector $\beta$ (for any choice of $P$) and that its covariance matrix is $(X^\top PX)^{-1}X^\top P\Sigma PX(X^\top PX)^{-1}$.

(e) What is the distribution of $\hat{\beta}$?

(f) Clearly, $\hat{\beta}$ depends on the choice of the matrix $P$. A best choice would be such that the covariance matrix is 'minimal'. This best choice is obtained for $P = \Sigma^{-1}$. What (rather obvious) additional assumption would (in principle) be needed to effectively compute $\hat{\beta}$? What is the resulting distribution for $\hat{\beta}$ in this case?

(g) In the classical set up, the elements of the vector $\varepsilon$ are IID normal random variables. Show that in the classical set up with the best choice of $P$, the estimator $\hat{\beta}$ doesn't depend on $P$ anymore. The additional assumption under (e) is not needed anymore to compute $\hat{\beta}$. But for testing of the $\beta$ parameters it could still be useful. What is the consequence in this context of imposing or not the additional assumption? What are the test statistics and their distributions that should be used?

72. Consider an IID sample $X_1, \ldots, X_n$, where the $X_k$ are nonnegative, having a density $f_\theta(x) = c(\theta)x\exp(-\frac{1}{2}\theta x^2)$ for $x \geq 0$ (and zero for $x < 0$) and $\theta > 0$ an unknown parameter.

(a) Show that $c(\theta) = \theta$.

(b) Show that for $p > -2$ it holds that $\mathbb{E}_\theta X_1^p = (\frac{2}{\theta})^{p/2}\Gamma(1 + \frac{1}{2}p)$. *Hint: In the integral that you need, you perform the substitution* $x = \sqrt{\frac{2}{\theta}}\sqrt{y}$ *after which you use the definition of the Gamma function.*

(c) Show that, for a fixed parameter $\theta$, the random variable $Y := X_1\sqrt{\theta}$ has density equal to $g(y) = y\exp(-\frac{1}{2}y^2)$ for $y > 0$. Determine also the cumulative distribution function $G$, i.e. give a formula for $G(y)$ for all(!) $y \in \mathbb{R}$. [Note that both functions don't depend on $\theta$.]

(d) What is the distribution of $\theta X_1^2$? And what is the distribution of $\theta\sum_{i=1}^n X_i^2$?

(e) Compute, based on $X_1, \ldots, X_n$, the maximum likelihood estimator of $\theta$, call it $\hat{\theta}_n$.

(f) Compute the moment estimator using the average of the $p$-th powers, $A_n^p := \frac{1}{n}\sum_{i=1}^n X_i^p$ ($p > -2$). Show that this is a consistent estimator of $\theta$.

(g) For each $p$ there is a corresponding moment estimator. Which one among them should be preferred? (*No complicated computations please!*)

(h) Test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 = \theta_1$, where $\theta_0 > \theta_1$, at a certain significance level $\alpha$. Use the Neyman-Pearson test and show that the test is equivalent to rejecting $H_0$ for large values of $\sum_{i=1}^n X_i^2$, $\sum_{i=1}^n X_i^2 > c(\alpha)/\theta_0$ say, where $c(\alpha)$ is an appropriate constant depending on $\alpha$. Describe $c(\alpha)$.

(i) Is the test of the previous question UMP for testing $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta < \theta_0$ (at the same significance level). Same question for the alternative $H_1 : \theta > \theta_0$.

(j) Consider the testing problem $H_0 : \theta \geq \theta_0$ and $H_1 : \theta < \theta_0$, using the test as in (h). Show that $\sup_{\theta \geq \theta_0} \mathbb{P}_\theta(\text{Reject } H_0) = \alpha$, i.e. the significance level of the test still equals $\alpha$.

73. Let $X$ have an exponential distribution with parameter $\lambda > 0$, so $X$ has density $f(x) = \lambda \exp(-\lambda x)$ for $x \geq 0$. Let $U$ be *independent* of $X$ such that $\mathbb{P}(U = +1) = \mathbb{P}(U = -1) = \frac{1}{2}$, and put $Z = UX$.

(a) We are interested in the distribution function $F_Z$ of $Z$. Show, split the event $\{Z \leq z\}$ into two sub-events according to $U = \pm 1$, that

$$F_Z(z) = \begin{cases} 1 - \frac{1}{2}\exp(-\lambda z) & \text{if } z \geq 0 \\ \frac{1}{2}\exp(\lambda z) & \text{if } z < 0. \end{cases}$$

(b) Show that the density $f_Z$ of $Z$ is given by $f_Z(z) = \frac{1}{2}\lambda \exp(-\lambda |z|)$ for $z \in \mathbb{R}$.

(c) Give a rough (but informative) sketch of the graph of $f_Z$. Thereby you pay attention to the values of $f_Z$ for $z$ near zero and for $z \to \pm\infty$. You may choose a value of $\lambda$ at will, should you find that convenient.

(d) From the previous item you can immediately deduce what $\mathbb{E}\, Z$ is. How? Verify your answer by using the definition of $Z$.

(e) Show that $\mathbb{V}\mathrm{ar}\,(Z) = \frac{2}{\lambda^2}$. *Hint: use the definition of $Z$.*

In the remainder of this exercise we assume to have an IID sample $Z_1, \ldots, Z_n$, each of the $Z_i$ having the density as in (b), which constitute our observations.

(f) Give the expression for the joint density of the vector $(Z_1, \ldots, Z_n)$ and show that the maximum likelihood estimator $\hat{\lambda}$ of $\lambda$ is given by $\hat{\lambda} = n / \sum_{i=1}^n |Z_i|$.

(g) Compute the Fisher information $I(\lambda)$ in a single observation $Z_1$.

(h) Give an approximate $(1-\alpha)$-confidence interval for $\lambda$ based on $\hat{\lambda}$. (Make sure that the limits of this interval don't depend on the unknown $\lambda$.)

74. Consider the linear regression model $Y = X\beta + e$, where $Y$ and $e$ are $2n$-dimensional random vectors, $X \in \mathbb{R}^{2n\times 2}$ and $\beta = (\beta_0, \beta_1)^\top$. Moreover, the design matrix $X$ is of the special form

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_n & \mathbf{1}_n \\ \mathbf{1}_n & \mathbf{0}_n \end{pmatrix},$$

where $\mathbf{1}_n$ is an $n$-dimensional column vector whose elements are all equal to 1 and $\mathbf{0}_n$ is an $n$-dimensional column vector whose elements are all equal to 0. We furthermore assume that the elements of the vector $e$ are *iid* random variables with zero mean and variance $\sigma^2$. Let $\overline{Y}_1 = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\overline{Y}_2 = \frac{1}{n}\sum_{j=n+1}^{2n} Y_j$.

(a) Show that the least squares estimator of $\beta$ is given by

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \overline{Y}_2 \\ \overline{Y}_1 - \overline{Y}_2 \end{pmatrix}.$$

*It may be helpful (not necessary) to show first* $(X^\top X)^{-1} = \frac{1}{n}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.

(b) Let $\hat{e}$ be the vector of residuals, $\hat{e} = Y - X\hat{\beta}$. Show that $\hat{e}^\top\hat{e} = \sum_{i=1}^n (Y_i - \overline{Y}_1)^2 + \sum_{j=n+1}^{2n} (Y_j - \overline{Y}_2)^2$.

(c) What is $\mathbb{E}\,(\hat{e}^\top\hat{e})$? Give an unbiased estimator $\widehat{\sigma^2}$ of $\sigma^2$.

Suppose a group of students is split into two subgroups of equal size. In the first subgroup students get additional training above the ordinary training, the second group follows the ordinary training only. This situation is reflected by the second column of the matrix $X$. The question is whether the extra training has a positive effect on the exam results of the students.

(d) Formulate a hypothesis testing problem in terms of the parameter vector $\beta$ that reflects the research issue. Give a suitable test statistic and describe the rejection region.

Assume that the usual normality assumptions on the error terms of $e$ and on the $Y_i$ are valid. Suppose that $n = 10$ and that in the first group the average exam result is 7.8 and 7.3 in the second group. Moreover, $\sum_{i=1}^n Y_i^2 = 610$ and $\sum_{i=n+1}^{2n} Y_i^2 = 534.50$.

(e) What is the result of the test at significance level $\alpha = 0.05$? Can we conclude from the observations that the training results in an advantageous effect on the performance of the students at the exam? Same questions for the case $\alpha = 0.01$.