Solutions to exam-0-4

Problem 1 (20 points)
You are given the following five unadjusted p-values for testing $H_1, \ldots, H_5$

$$0.99734, 0.60008, 0.13896, 0.00773, 0.00097,$$

where 0.99734 is the p-value for H_1, 0.60008 is the p-value for H_2 and so on. Calculate Bonferroni, Holm, and Benjamini and Hochberg adjusted p-values. For each of the three methods decide which hypotheses we reject if we use $\alpha = 0.05$.

Solution: Here, we obtain Bonferroni adjusted p-values by multiplying the above p-values by 5. Results larger than 1 are set equal to 1. We find

$$1.00; 1.00; 0.6948; 0.03865; 0.00485.$$

Hence we reject H_4 and H_5, because their Bonferroni adjusted p-values are both less than 0.05.
To find Holm adjusted p-values we first order them. Here the smallest is multiplied by 5 the second smallest by 4, the third smallest by 3 and so on. Then the adjusted p-values increasingly ordered are (again using the convention that p-values bigger than 1 are set equal to 1)

$$0.00485; 0.03092; 0.41688; 1.00; 1.00.$$

Hence, we reject H_4 and H_5 because they correspond to the second smallest and smallest ordered p-values.
For Benjamini and Hochberg we multiply the smallest p-value by 5, the second smallest by 5/2, the third smallest by 5/3 and so on. Then we obtain

$$0.00485; 0.019325; 0.231600; 0.750100; 0.997340.$$

We need to find the largest of these which is less than 0.05. Here this is $p_{(2)}$ and we reject all hypotheses with p-value less or equal than $p_{(2)}$. Here this means we reject H_4 and H_5.

Problem 2 (10 + 10 points)
Assume that the cumulative distribution function of the random variable X is given by

$$F_\lambda(x) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^2\right),\ x > 0,\ \text{and zero otherwise,}$$

where $\lambda > 0$. For testing

$$H: \lambda^2 \leq 2, \quad A: \lambda^2 > 2,$$

we use, based on a sample X of size 1, the test statistic $T(X) = X^2$ and we reject H at level α if X exceeds the critical value $c(\alpha)$.

(i) Find the critical value $c(\alpha)$ if we test at $\alpha = 0.05$;

(ii) Calculate the power of the test at $\lambda = 4$.

Solution: (i) We need to find a real number $c(0.05)$ such that

$$\mathbb{P}_{\lambda^2=2}(X^2 \le c(0.05)) = 0.95.$$

This is equivalent to (note that $\lambda^2 \le 2$ is the same as $\lambda \le \sqrt{2}$)

$$\mathbb{P}_{\lambda^2=2}(X \le \sqrt{c(0.05)}) = 1 - \exp\left(-\left(\frac{\sqrt{c(0.05)}}{\sqrt{2}}\right)^2\right) = 1 - \exp\left(-\frac{c(0.05)}{2}\right) \stackrel{!}{=} 0.95.$$

Here $\stackrel{!}{=}$ means 'must be equal to'. Solving for $c(0.05)$ we find

$$c(0.05) = -2\log(0.05) = 5.991465.$$

(ii) We need to calculate

$$\mathbb{P}_{\lambda=4}(X^2 > 5.991465) = 1 - \mathbb{P}_{\lambda=4}(X^2 \le 5.991465) = 1 - \mathbb{P}_{\lambda=4}(X \le \sqrt{5.991465}),$$

which equals

$$1 - \left[1 - \exp\left(-\left(\frac{\sqrt{5.991465}}{4}\right)^2\right)\right] = 0.687656.$$

Remarks: The distribution is not an exponential distribution nor a gamma distribution. It is a Weibull distribution with shape parameter being equal to 2.
There is no need to note that '$\lambda^2 \le 2$ is the same as $\lambda \le \sqrt{2}$' because you can directly work with λ^2 as this term appears in the cdf.

Problem 3 (10 points)
For $0 < p < 1$ consider the following probability distribution

$$\mathbb{P}(Y = y) = \frac{(1-p)^{y-1}p}{1-(1-p)^{10}}, \text{ for } y = 1, \dots, 10,$$

which takes only the values $1, 2, \dots, 10$. In other words the probability mass function g_Y^p of this random variable is

$$g_Y^p(y) = \frac{(1-p)^{y-1}p}{1-(1-p)^{10}}, \text{ for } y = 1, \dots, 10.$$

In class (Lecture 6) we discussed a particular form for probability mass functions given by

$$f_\theta^Y(y) = \exp\left(\frac{y\theta - b(\theta)}{\psi} - c(\psi, y)\right), \; y \in D,$$

where $\theta \in \Theta$ and ψ are real-valued parameters, D is the support of the distribution of Y, and b and c are real-valued functions. Is it possible to write g_Y^p in this form?
Solution: We rewrite g_Y^p as

$$\begin{aligned}
&\exp\left(\log\left(\frac{(1-p)^{y-1}p}{1-(1-p)^{10}}\right)\right) \\
=&\exp\left((y-1)\log(1-p) + \log(p) - \log(1-(1-p)^{10})\right) \\
=&\exp\left(y\log(1-p) - \log(1-p) + \log(p) - \log(1-(1-p)^{10})\right).
\end{aligned}$$

Comparing this with f_θ^Y we see that we must have

$$\theta = \log(1-p), \psi \equiv 1, c(\psi, y) \equiv 0$$

and using that $p = 1 - \exp(\theta)$

$$b(\theta) = \theta - \log(1 - \exp(\theta)) + \log(1 - \exp(\theta)^{10}).$$

Remarks: Note that ψ and $c(\psi; y)$ are not allowed to depend on θ or which is the same on $\log(1-p)$. Note also that b must be given as a function of θ.

Problem 4 (7.5+7.5 points)
In class we related the expectation of a random variable Y to a linear function $\sum_{j=1}^d \beta_j x_j$ using a link function h. Assume that the distribution of Y is given by

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

where we have for the parameter p that $0 < p \leq 1$. This implies that the expectation of $\mathbb{E}[Y]$ equals

$$\mathbb{E}[Y] = \frac{1}{p}.$$

For each of the following alternative choices of the link function h, argue if it is meaningful to use them to relate $\mathbb{E}[Y]$ and $\sum_{j=1}^d \beta_j x_j$ by $\mathbb{E}[Y] = h(\sum_{j=1}^d \beta_j x_j)$. Explain your answer.

(i) $h_1(x) = |x| + 1$, $x \in \mathbb{R}$;

(ii) $h_2(x) = \int_0^{|x|} y^2\, dy$, $x \in \mathbb{R}$.

Solution: Note first that p is an element of $(0, 1]$ which implies that $\mathbb{E}[Y]$ is an element of $[1, \infty)$. Therefore any meaningful link function must also map to $[1, \infty)$ (or to $(1, \infty)$).
(i) is meaningful as its range is $[1, \infty)$.
(ii) is not meaningful because the integral equals $(1/3)|x|^3$ for any $x \in \mathbb{R}$. Hence, the range of the link function is $[0, \infty)$.
Remark: This is not a binomial distribution. It is a geometric distribution.

Problem 5 (7.5+7.5 points)
Assume our data come from the linear model

$$Y_i = \sum_{j=1}^{60} \beta_j X_{ij} + \epsilon_i,\ i = 1, \ldots, 10,$$

with ϵ_i, $1 \leq i \leq 10$, independent and normally distributed with expectation zero and variance σ^2. Unfortunately the observations $y_1, \ldots, y_{10}$, and $x_{11}, \ldots, x_{10\,60}$ were lost. What is known is that the X_{ij} were independent and each normally distributed with expectation 2 and variance 10. A friend of you tells you that he was additionally given the following two vectors

(i) $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_{60})$ with $\bar{\beta}_j = 1 + j$ for $j = 1, \ldots, 12$ and $\bar{\beta}_j = 0$ otherwise;

(ii) $\breve{\boldsymbol{\beta}} = (\breve{\beta}_1, \ldots, \breve{\beta}_{60})$ with $\breve{\beta}_j = 1 + j$ for $j = 1, \ldots, 5$, $\breve{\beta}_{59} = 2.5$ and $\breve{\beta}_j = 0$ otherwise.

Given this information only decide for both $\bar{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\beta}}$ whether they could potentially be the solution to the following minimization problem

$$\text{minimize w.r.t. } \boldsymbol{\beta} : \ \frac{1}{10} \sum_{i=1}^{10} \left(y_i - \sum_{j=1}^{60} \beta_j \, x_{ij} \right)^2 + 3 \sum_{j=1}^{60} |\beta_j|,$$

where $y_1, \ldots, y_{10}$ and $x_{11}, \ldots, x_{10\,60}$ are the unknown observations.
Explain your answers briefly.

Solution: From class we know that in case of continuous regressors the LASSO estimator is unique and has at most $\min\{n, d\}$ non-zero entries. This result can be applied here because the regressors are independent and each has a continuous distribution which implies that their joint distribution is also a continuous distribution. Then the vector in (i) can be ruled out as it has more than 10 non-zero entries. The estimator in (ii) cannot be ruled out and is therefore a potential solution.

Problem 6 (10 points)
We discussed in class that there can be multiple solutions to the (LASSO) minimization problem

$$\text{minimize w.r.t. } \boldsymbol{\beta} : \ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{d} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{d} |\beta_j|.$$

Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ be two different solutions for this minimization problem. Prove or disprove that $X\hat{\boldsymbol{\beta}}_1 = X\hat{\boldsymbol{\beta}}_2$ where, as usual, X is the design matrix.
Hints: $X\hat{\boldsymbol{\beta}}_1$ and $X\hat{\boldsymbol{\beta}}_2$ are in $\mathbb{R}^n$, the mapping $z \to ||y - z||_2^2$ is strictly ...
Solution: We rewrite and use convexity of $x \to |x|$ to find for $0 \le \alpha \le 1$

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \left(\sum_{j=1}^{d} \alpha \hat{\beta}_{1j} x_{ij} + (1-\alpha) \sum_{j=1}^{d} \hat{\beta}_{2j} x_{ij} \right) \right)^2 + \lambda \sum_{j=1}^{d} |\alpha \hat{\beta}_{1j} + (1-\alpha) \hat{\beta}_{2j}| \\
&\le \frac{1}{n} ||y - \alpha X \hat{\boldsymbol{\beta}}_1 - (1-\alpha) X \hat{\boldsymbol{\beta}}_2||_2^2 + \lambda \sum_{j=1}^{d} \alpha |\hat{\beta}_{1j}| + \lambda \sum_{j=1}^{d} (1-\alpha) |\hat{\beta}_{2j}|,
\end{aligned}$$

where as usual $|| \cdot ||_2$ denotes the Euclidean scalar product. The mapping $z \to ||y - z||_2^2$ where $z \in \mathbb{R}^n$ is strictly convex; cf. hints. Hence,

$$\begin{aligned}
&\frac{1}{n} ||y - \alpha X \hat{\boldsymbol{\beta}}_1 - (1-\alpha) X \hat{\boldsymbol{\beta}}_2||_2^2 + \lambda \sum_{j=1}^{d} \alpha |\hat{\beta}_{1j}| + \lambda \sum_{j=1}^{d} (1-\alpha) |\hat{\beta}_{2j}| \\
&< \frac{1}{n} \alpha ||y - X \hat{\boldsymbol{\beta}}_1||_2^2 + (1-\alpha) ||y - X \hat{\boldsymbol{\beta}}_2||_2^2 + \lambda \sum_{j=1}^{d} \alpha |\hat{\beta}_{1j}| + \lambda \sum_{j=1}^{d} (1-\alpha) |\hat{\beta}_{2j}| \\
&= \alpha \left(\frac{1}{n} ||y - X \hat{\boldsymbol{\beta}}_1||_2^2 + \lambda \sum_{j=1}^{d} |\hat{\beta}_{1j}| \right) + (1-\alpha) \left(||y - X \hat{\boldsymbol{\beta}}_2||_2^2 + \lambda \sum_{j=1}^{d} |\hat{\beta}_{2j}| \right),
\end{aligned}$$

where $<$ is due to the strict convexity. The last line equals the minimum of the function

$$\boldsymbol{\beta} :\to \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{d}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{d}|\beta_j|, \tag{1}$$

because $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ both minimize this function (and α and $(1-\alpha)$ add up to 1). This gives a contradiction because according to the above the function value at $\alpha\hat{\boldsymbol{\beta}}_1 + (1-\alpha)\hat{\boldsymbol{\beta}}_2$ would be smaller.
Remark: Note the difference to what we discussed in class. There we noticed that the function in Equation (1) is convex but not strictly convex if $d > n$. Yet, here we argue about $X\hat{\boldsymbol{\beta}}_1$ and $X\hat{\boldsymbol{\beta}}_2$ which are in $\mathbb{R}^n$ and as said in the hint $z \to ||y-z||_2^2$ is strictly convex for $z \in \mathbb{R}^n$.

Problem 7 (10 points)
For $k = 1, \ldots, 5$ let I_k be a confidence interval for θ_k, $k = 1, \ldots, 5$, with coverage probability $1-\alpha$. Assuming that the I_k are independent how do we need to choose α such that $I_1 \times \ldots \times I_5$ is a simultaneous confidence interval for $(\theta_1, \ldots, \theta_5)$ at level 95% (, i.e. $\mathbb{P}((\theta_1, \ldots, \theta_5) \in I_1 \times \ldots \times I_5 = 0.95)$)?
Solution: We need to find α such that

$$(1-\alpha)^5 = 0.95.$$

This is equivalent to $\alpha = 1 - 0.95^{\frac{1}{5}} = 0.01020622$.