Statistics 2 Unit 1

Team 8

April 17, 2018

Contents

1 Task 61

1.1 a)

The posterior :

$$f_{\Theta|T}(\theta|t) \propto L_{T|\Theta}(t|\theta)f_{\Theta}(\theta)$$

Since the t_i are uniformly distributed, the likelihood function:

$$L_{T|\Theta}(t|\theta) = f(t|\theta) = \prod_{i=1}^{n} f(t_i|\theta) = \frac{1}{\theta^n}$$

Considering the prior function, it is Pareto distributed with paramters α, β:

$$f_{\Theta}(\theta) = f(\theta|\alpha, \beta) \propto \theta^{-(\alpha+1)}$$

$\Rightarrow$ the posterior is:

$$f_{\Theta|T}(\theta|t) \propto L_{T|\Theta}(t|\theta)f_{\Theta}(\theta) = \frac{1}{\theta^n}\theta^{-(\alpha+1)}$$

It is also Pareto-distributed
Plugging in: $\alpha = 7, \beta = 4, t_{max} = 14, n = 5$ we have that $t|\theta$ are distributed as Pareto with parameters 12 and 14.

1.2 b)

$$\mathbb{E}\left[[\theta|t]\right] = \frac{\alpha\beta}{\alpha - 1} = \frac{12 * 14}{12 - 1} = 15.27$$

1.3 c)

$$\mathbb{P}\left(\beta < \theta < C|t\right) = 1 - a$$

The critical value C for the upper bound in the interval θ must be larger than the posterior $\beta = 14$, otherwise the density will be 0.
Using the Pareto distribution we have:

$$F(\theta|\alpha, \beta) = 1 - (\frac{\beta}{\theta})^{\alpha}$$
$$\mathbb{P}\left(\beta < \theta < C|t\right) = 1 - a$$

Finally:

$$C = \frac{\beta}{(1-a)^{\frac{1}{\alpha}}} = \frac{14}{(0.05)^{\frac{1}{12}}} = 18$$

And the 95% HPD is (14, 18).

1.4 d)

If $\mathbb{P}\left(H_0|t\right) < 0.5$ then we reject the null hypothesis. So:

$$\mathbb{P}\left(H_0|t\right) = \mathbb{P}\left(0 < \theta < 15|t\right) - \mathbb{P}\left(14 < \theta < 15|t\right) =$$
$$= F(15|\alpha = 12, \beta = 14) - F(14|\alpha = 12, \beta = 14) = 0.56 > 0.5$$
$$\Rightarrow \text{ we can't reject the null hypothesis.}$$

2 Task 62

Let $X_1, ..., X_n$ be iid from a geometric distribution, i.e. $f(x) = p(1-p)^x, x = 0, 1, ...$.

2.1 a)

The likelihood function is given by:

$$l(p, x_i) = \prod_{i=0}^{n} p(1-p)^{x_i} = p^n(1-p)^{\sum x_i}.$$

Let's derive the posterior where the prior is a beta distribution:

$$f(p|x_i) \propto l(p, x_i)\, f(p) = p^n(1-p)^{\sum x_i}\, \frac{p^{\alpha-1}\,(1-p)^{\beta-1}}{B(\alpha, \beta)} = p^{\alpha+n-1}\,(1-p)^{\beta+\sum x_i-1}.$$

Thus we see that the posterior is again a beta distribution with parameters $(\alpha + n, \beta + \sum x_i)$. Therefore beta is a conjugate prior.

The Bayes estimator is defined as the expected value of the posterior distribution. So we get the Bayes estimator by applying the formula for the expected value of a beta distribution:

$$\mathbb{E}\left[p|x_i\right] = \frac{\alpha + n}{\alpha + \beta + \sum x_i + n}$$

2.2   b)

The Jeffrey prior is defined as $\pi(p) \propto \sqrt{I(p)}$.
So let's find the Fisher information.

$$log(f(x)) = log(p) + xlog(1-p)$$

$$\frac{\partial log(f(x))}{\partial p} = \frac{1}{p} - \frac{x}{1-p}$$

$$\frac{\partial^2 log(f)}{\partial p^2} = -\frac{1}{p^2} - \frac{x}{(1-p)^2}$$

Then the Fisher information is:

$$I(p) = -\mathbb{E}\left[\frac{\partial^2 log(f)}{\partial p^2}\right] = \frac{1}{p^2} + \mathbb{E}\left[\frac{x}{(1-p)^2}\right] = \frac{1}{p^2} + \frac{\frac{1-p}{p}}{(1-p)^2} = \frac{1}{p^2(1-p)}$$

$$\pi(p) \propto \sqrt{I(p)} = \frac{1}{p\sqrt{(1-p)}}.$$

The posterior is then found by

$$f(p|x_i) \propto l(x_i)\, \pi(p) = p^n(1-p)^{\sum x_i} \frac{1}{p\sqrt{(1-p)}} = p^{n-1}\,(1-p)^{\sum x_i - 1/2}.$$

We see that the posterior is again beta distributed with parameters $(n, \sum_{i=0}^{n} x_i + 1/2)$.

Bayes estimator: $\mathbb{E}[p|x_i] = \frac{n}{n+\sum x_i + 1/2}$

3   Task 63

3.1   a:

We have 1 unknown parameter θ, thus we need to calculate the fisrt moment $E(X)$. For the discrete case:

$E(X) = \sum_{i=1}^{n} p_i X_i.$

$E(X) = 0*2\theta/3 + 1*\theta/3 + 2*2(1-\theta)/3 + 3*(1-\theta)/3 = (\theta + 4 - 4\theta + 3 - 3\theta)/3 = (-6\theta + 7)/3$

4

$=> \theta = -(3E(X) - 7)/6$

We can estimate $E(X)$ using the sample from the task:

$\hat{\mu} = 3 * 2/10 + 0 * 2/10 + 2 * 3/10 + 1 * 3/10 = 3/2$

$=> \hat{\theta}_{MoM} = -(3 * 3/2 - 7)/6 = 5/12$

3.2   b:

We need to find standard deviation of $\hat{\theta}_{MoM}$

$$Var(\hat{\theta}_{MoM}) = Var(-3\hat{\mu}/6 + 7/6) = 0.5^2 Var(\hat{\mu})$$

we will use the knowledge that:

$$Var(\hat{\mu}) = Var(X)/n$$

$$Var(X) = \sum p_i x_i^2 - E(X)^2 = \frac{\theta + 8(1 - \theta) + 9(1 - \theta)}{3} - 9/4 = 1.19$$

For the $\hat{\theta} = 5/12$.

$$Sd(\hat{\theta}_{MoM}) = 0.5\sqrt{1.19/10} = 0.1725$$

3.3   c:

The likelihood function for the discret case will be:

$$L(\theta) = \prod_{i=1}^{n} P(X = x_i),$$

where $P(X = x_i)$ is the probability mass function.

$$L(\theta) = 2\theta/3 * \theta/3 * 2(1 - \theta)/3 * (1 - \theta)/3$$

$$l(\theta) = \log(2\theta/3) + \log(\theta/3) + \log(2(1 - \theta)/3) + \log((1 - \theta)/3)$$

$$\frac{\partial l(\theta)}{\partial \theta} = 1/\theta + 1/\theta - 1/(1 - \theta) - 1/(1 - \theta) = \frac{2 - 4\theta}{\theta - \theta^2} = 0$$

$=> \hat{\theta}_{MLE} = 0.5$

3.4 d:

Recall the theorem that the maximum likelihood estimator $\hat{\theta}$ is asymptotically distributed as $N(\theta_0, \frac{1}{nI_\theta})$.

$$I_\theta = -E_\theta\left(\frac{\partial l^2}{\partial \theta^2}\right) = -E_\theta\left(\frac{-4(\theta - \theta^2) - (1 - 2\theta)(2 - 4\theta)}{(\theta - \theta^2)^2}\right) = \frac{4\theta^2 - 4\theta + 2}{(\theta - \theta^2)^2}$$

We can use our estimate for θ and calculate the Fisher Information:

$I = \frac{4*0.5^2 - 4*0.5 + 2}{(0.5 - 0.5^2)^2} = \frac{1 - 2 + 2}{0.125^2} = 1/0.5^2$

$$Var(\hat{\theta}) = \frac{1}{nI} = \frac{0.5^2}{10}$$

$=> sd(\hat{\theta}) = 0.5/\sqrt{10} = 0.1581$

3.5 e:

The joint distribution of X and θ:

$$f_{X,\theta}(x, \theta) = f_{X|\theta}(x|\theta)f_\theta(\theta)$$

We know that prior $\sim U(0, 1)$. Hence, $f_\theta(\theta) = \frac{1}{b-a} = 1$
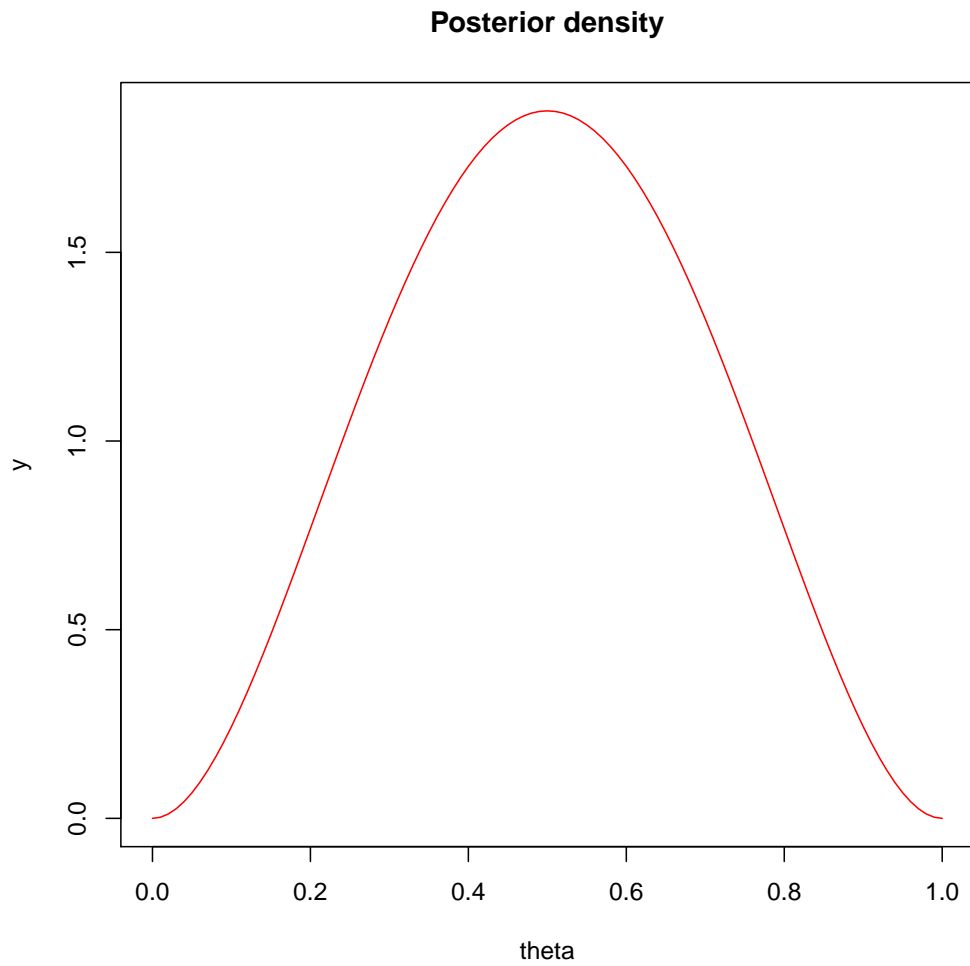
$=> f_{X,\theta}(x, \theta) = f_X|\theta(x|\theta)$

It gives us, that joint density is equal to likelihood. Hence:

$$f_{X,\theta}(x, \theta) = \prod_{i=1}^{n} P(X = x_i) = 2\theta * \theta * 2 * (1-\theta)(1-\theta)/81 = (4\theta^4 - 8\theta^3 + 4\theta^2)/81$$

We also know that posterior is equal to:

$$f_{\theta|X}(\theta|x) = \frac{f_{x|\theta}(x|\theta)f_\theta(\theta)}{\int f_{x|\theta}(x|\theta)f_\theta(\theta)d\theta} = \frac{f_{x|\theta}(x|\theta)}{\int_0^1 f_{x|\theta}(x|\theta)d\theta} = \frac{(4\theta^4 - 8\theta^3 + 4\theta^2)/81}{\int_0^1 (4\theta^4 - 8\theta^3 + 4\theta^2)/81 d\theta} =$$

$$= \frac{(4\theta^4 - 8\theta^3 + 4\theta^2)/81}{2/1215} = 7.5(4\theta^4 - 8\theta^3 + 4\theta^2) = 30\theta^4 - 60\theta^3 + 30\theta^2$$

Posterior density

On the graph we see that mode is equal to 0,5, i.e. mode is $\hat{\theta}_{MLE}$, because in order to find $\hat{\theta}_{MLE}$ we needed to find maximum of likelihood function and in this case posterior density is proportional to the likelihood (it follows from the fact that prior $\sim U(0,1)$).

4 Task 65

4.1 a)

Suppose μ is known. Then, the loglikelihood function is:

$$\ell\left(\sigma\right) = \log\left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_i - \mu)^2}{2\sigma^2}}\right] = -n\log\left(\sigma\right) - \frac{n}{2}\log\left(2\pi\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

Differentiation of $\ell\left(\sigma\right)$ wrt σ,

$$\ell'\left(\sigma\right) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - \mu)^2$$

The MLE $\widehat{\sigma}$ of σ is obtained by solving $\ell'\left(\sigma\right) = 0$
Thus,

$$\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2}$$

4.2 b)

Suppose σ is known. The loglikelihood function is the same:

$$\ell\left(\mu\right) = \log\left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_i - \mu)^2}{2\sigma^2}}\right] = -n\log\left(\sigma\right) - \frac{n}{2}\log\left(2\pi\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

Differentiating $\ell(\mu)$ wrt μ, we have

$$\ell'(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)$$

The MLE $\widehat{\mu}$ of μ is obtained by solving for μ which satisfies the equation

$$\ell'(\mu) = 0$$

Thus,

$$\widehat{\mu} = \overline{X}$$

4.3   c)

Note that $\overline{X}$ is an unbiased estimate for μ since $E\left(\overline{X}\right) = \mu$

Now, calculate $I\left(\mu\right)$

$$I\left(\mu\right) = -E\left[\frac{\partial^2}{\partial\mu^2}\log f\left(X_1|\,\mu\right)\right] =$$

$$= -E\left[\frac{\partial^2}{\partial\mu^2}\left(-\log\left(2\sigma\right) - \frac{1}{2}\log\left(2\pi\right) - \frac{\left(X_1 - \mu\right)^2}{2\sigma^2}\right)\right]$$

$$= \frac{1}{\sigma^2}$$

$$\Rightarrow Var\left(\overline{X}\right) = \frac{\sigma^2}{n} = \frac{1}{nI\left(\mu\right)}$$

Cramer-Rao Inequality states that if $T\left(X_1, \ldots, X_n\right)$ is any unbiased estimate of μ, then

$$Var\left(T\right) \geq \frac{1}{nI\left(\mu\right)} = Var\left(\overline{X}\right)$$

This imlies that no other unbiased estimate of μ can have a smaller variance than $\overline{X}$.

5   Task 68

Let $X_1, ..., X_n$ iid $\sim P(\lambda)$ and let T be a test statistics such that $T = \sum_{i=1}^{n} X_i$, which we will further denote as $\sum X_i$ for better readability.

a)

Let's check if the given test statistics is sufficient. We have the following given:

$$P(X_1, ..., X_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{\sum x_i}}{x_1!...x_n!}$$

$$P(T = t) = e^{-n\lambda}\frac{(n\lambda)^t}{t!}.$$

Then we can look on its conditional distribution in order to check whether the test statistics contains all the necessary information. The joint density is

$$P(X_1, ..., X_n, T = t) = e^{-n\lambda}\frac{\lambda^{\sum x_i}}{x_1!...x_n!}$$

if $T = \sum_{i=1}^{n} X_i$ and 0 else. So we need only consider the former case. Then the conditional is

$$P(X_1, ..., X_n | T = t) = \frac{P(X_1, ..., X_n, T = t)}{P(T = t)} = \frac{e^{-n\lambda} \frac{\lambda^{\sum x_i}}{x_1!...x_n!}}{e^{-n\lambda} \frac{n^t \lambda^t}{t!}} \overset{T=t}{=} \frac{t!}{x_1!, ..., x_n!} \frac{1}{n^{\sum x_i}}$$

$$= \frac{(\sum_{i=1}^{n} X_i)!}{x_1!, ..., x_n!} \frac{1}{n^{x_1}} \frac{1}{n^{x_2}} ... \frac{1}{n^{x_n}}.$$

It is sufficient, since in the multinomially distributed conditional is no longer a parameter λ contained. This means all the necessary information are contained in the necessary statistic for parameter estimation if $T = t$.

b)

Now we check the same for the case where $T = X_1$.

$$P(X_i | T = X_1) = \frac{P(X_1, ..., X_n, T = X_1)}{P(T = X_1)} = \frac{e^{-n\lambda} \frac{\lambda^{\sum x_i}}{x_1!...x_n!}}{e^{-\lambda} \frac{\lambda^t}{t!}} = \frac{1}{x_2!, ..., x_n!} \frac{e^{-(n-1)\lambda} \lambda^{\sum x_i}}{\lambda^{x_1}}.$$

We see that our likelihood function conditioned on the test statistics $T = X_1$ doesn't eliminate parameter. Therefore, if we have this test statistics, not all information is included. Thus, this test statistics is not sufficient.

c)

For using the factorization theorem we have to have a look closer on the likelihood function:

$$P(X_1, ..., X_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum x_i} \frac{1}{x_1!...x_n!}$$

where we have $g(\sum x_i, \lambda) = e^{-n\lambda} \lambda^{\sum x_i}$ and $h(x_i) = \frac{1}{x_1!...x_n!}$.
We can split the likelihood in two functions where in one is the parameter included and in the other not. This fulfills the factorization theorem. This is sufficient and necessary for a showing that the test statistics $T = \sum_{i=1}^{n} X_i$ is sufficient.

6 Task 69

We will consider 3 cases, where α is shape parameter and β is rate parameter:

1) α, β are unknown;
2) α is known, β is unknown:
3) α is unknown, β is known.

First of all, we recall the factorization theorem:

A necessary and sufficient condition for $T(X_1, \ldots, X_n)$ to be sufficient for a parameter θ is that the joint probability function factors in the form:

$$f_\theta(x_1, \ldots, x_n) = h(x_1, \ldots, x_n) \, g_\theta(T(x_1, \ldots, x_n))$$

Now, we will consider the 1st case, when both parameters are unknown.
We have $X_1, \ldots, X_n$ which are i.i.d, so we can write the joint density ($X^n = X_1, \ldots, X_n$) as a product of densities:

$$f_{X^n}(x^n) = \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} = (\frac{\beta^\alpha}{\Gamma(\alpha)})^n \prod_{i=1}^n x_i^{\alpha-1} e^{-\beta \sum_{i=1}^n x_i}$$

Here we can notice that joint density satisfies the factorization theorem, where $h(x^n) = 1$ and joint density itself is $g_\theta(T(x_1, \ldots, x_n))$, because it is a function which is dependent on both parameters and on X through the sufficient statistics: $T(X^n) = (\prod_{i=1}^n x_i, \sum_{i=1}^n x_i)$. So, in the case when α, β are unknown, we have that sufficient statistics $T(X^n)$ is a two-dimensional.

Now let's consider the case when α is known, β is unknown:
We have the same joint density, but as far as we know α:

$$h(x^n) = \frac{\prod_{i=1}^n x_i^{\alpha-1}}{\Gamma(\alpha)^n}; \;\; g_\theta(T(x^n)) = \beta^{\alpha n} e^{-\beta \sum_{i=1}^n x_i}$$

In this case, $T(x^n) = \sum_{i=1}^n x_i$ will be sufficient statistics.

And in the last case, when α is unknown, β is known, we again need to group the joint density taking into account that that β is some constant:

$$h(x^n) = e^{-\beta \sum_{i=1}^n x_i}; \;\; g_\theta(T(x^n)) = (\frac{\beta^\alpha}{\Gamma(\alpha)})^n \prod_{i=1}^n x_i^{\alpha-1}$$

Here it's clear that when we know β, sufficient statistics is $T(x^n) = \prod_{i=1}^n x_i^{\alpha-1}$.

X	H_0	H_A	Λ	$P(\Lambda(x) < \Lambda(x_i)|H_0)$
x_3	0.3	0.1	3	0.7
x_1	0.2	0.1	2	0.5
x_2	0.3	0.4	0.75	0.2
x_4	0.2	0.4	0.5	0.0

7   Task 76

a)

We know, that $\Lambda = \frac{H_0}{H_A}$, thus:

$$\Lambda(x_1) = \frac{0.2}{0.1} = 2$$

$$\Lambda(x_2) = \frac{0.3}{0.4} = 0.75$$

$$\Lambda(x_3) = \frac{0.3}{0.1} = 3$$

$$\Lambda(x_4) = \frac{0.2}{0.4} = 0.5$$

thus the order from the highest Λ to the lowest would be x_3, x_1, x_2, x_4

b)

Let's rearrange the table, ordering the rows by decreasing Λ. We also add a column with $P(\Lambda(X) < \Lambda(x_i)|H_0)$, which is exactly the level of the test that rejects H_0 if $\Lambda(X) < \Lambda(x_i)$:

For $\alpha = 0.2$ the test that rejects H_0 is $\Lambda(X) < \Lambda(x_2)$ (or simply $\Lambda(X) < 0.75$).
For $\alpha = 0.5$ the test that rejects H_0 is $\Lambda(X) < \Lambda(x_1)$ (or simply $\Lambda(X) < 2$).

c)

From our prior probabilities we can compute the ratio of posterior probabilities by:

$$\frac{P(H_0|x)}{P(H_A|x)} = \frac{P(H_0)P(x|H_0)}{P(H_A)P(x|H_A)} > 1$$

Since $P(H_0) = P(H_A)$, we get

$$\frac{P(x|H_0)}{P(x|H_A)} > 1$$

We can see from part a) that only x_1 and x_3 favor H_0.

d)

From b) we can conclude that for $\alpha = 0.2$ the rejection region of the test is $\{x_4\} = \{x : \Lambda(X) < 3/4\}$.
For the posterior odds to be less than 1, the prior odds must be less than $4/3$, which corresponds to

$$P(H_0) = 1 - P(H_A) < \frac{4}{7}$$

For $\alpha = 0.5$ the rejection region of the test is $\{x_4, x_2\} = \{x : \Lambda(X) < 2\}$.
For the posterior odds to be less tan 1, the prior odds must be less than 0.5, which corresponds to

$$P(H_0) = 1 - P(H_A) < \frac{1}{3}$$

To include x_2 in the rejection region it must be that $P(H_0) > 4/7$. So for $\alpha = 0.5$ the prior probabilities are such that

$$4/7 < P(H_0) = 1 - P(H_A) < 1/3$$

8   Task 77

a) FALSE
Significance level is the probability of a type I error, i.e. the probability that H_0 is rejected when it is true.
b) FALSE
The decrease of the significance level means that the probability of type I error (rejection of the null if it is true) decreases. The power is the probability that the null hypothesis is rejected when is false. By the Neyman - Pearson lemma if the significance level decreases, so does the power.
c) FALSE

Same as a)

d) FALSE

The power is the probability that the null hypothesis is correctly rejected, not falsely.

e) FALSE

When the test statistic falls in the rejection region of the test, the test procedure rejects the null hypothesis. This may lead to a correct decision of the null hypothesis is false, or to an incorrect decision (the type I error) if the null hypothesis is true. We do not know which.

f) FALSE

Both error types are serious. In Neyman-Pearson testing, we limit the size of the type I error and find the test that minimizes the type II error subject to the constraint. By controlling the size of the type I error, it is given precedence in specifying tests. In real life the answer actually depends on what is taken as H_0 and what is taken is H_A. For example, if for some medical test the H_0 is "the person is healthy", accepting the null hypothesis when it is false (type II error), i.e. assuming the person to be healthy when he/she is actually ill, is more serious. However, if we consider a criminal court procedure outcome (the final sentence) as a (to some extent) random variable and take as H_0 "the person is innocent", the type I error (to find an innocent person guilty) is considered in modern law theory to be worse than to free from a charge a person who is actually guilty.

g) FALSE

The power of a test is the conditional probability of correctly rejecting the null hypothesis when the alternate hypothesis is true. It is a function which is computed for every distribution in the alternate hypothesis.

h) TRUE

As a function of random variables, the likelihood ratio is also a random variable.