Student ID:

<u>Final Examination: QUESTION BOOKLET</u>

Prof. William Fithian

Fall 2018

- Do *NOT* open this question booklet until you are told to do so.
- Write your Student ID number at the top of this page.
- Write your solutions in this booklet.
- No electronic devices are allowed during the exam.
- Be neat! If we can't read it, we can't grade it.
- You can treat any results from lecture or homework as "known," and use them in your work without rederiving them, but do make clear what result you're using. You do not need to explicitly check regularity conditions for the theorems from class that required them.
- For a multi-part problem, you may treat the results of previous parts as given (if you don't prove the result for part (a), you can still use it to solve part (b)).
- I have starred some parts which I believe are the most difficult, and which I expect most students won't necessarily be able to solve in the time allotted. They are generally not worth more points than the less difficult parts, so don't waste too much time on them until you're happy with your answers to the latter.
- Be careful to justify your reasoning and answers. We are primarily interested in your understanding of concepts, so show us what you know.
- Good luck!

1. A curved Gaussian family (20 points, 4 points / part). Some useful facts for this problem:

- Recall that the Gaussian density function for $Z \sim N(\mu, \sigma^2)$ is

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Suppose that

$$X_1, \ldots, X_n = \begin{pmatrix} X_{1,1} \\ X_{1,2} \end{pmatrix}, \ldots, \begin{pmatrix} X_{n,1} \\ X_{n,2} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N_2(\mu(\theta), I_2),$$

for $\theta \in \mathbb{R}$ and $\mu(\theta) = \binom{\theta}{\theta^2}$.

(a) Show that $T(X) = \sum_i X_i \in \mathbb{R}^2$ is a minimal sufficient statistic but is not complete sufficient.

(b) Find the Fisher information $J_n(\theta)$, i.e. the information about θ in the complete sample.

(c) Consider the score test of $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$. Give an explicit expression for both the test statistic and its rejection threshold, and show that the test achieves finite-sample control of the Type I error rate.

(d) Find the asymptotic distribution of $\hat{\mu}_2 = \hat{\theta}^2$, the MLE for the expectation of $X_{i,2}$, when $\theta \neq 0$. Compare its asymptotic relative efficiency to the "obvious" estimator $\frac{1}{n}\sum_{i=1}^n X_{i,2}$.

(e) (*) If $\theta = 0$, find the asymptotic distribution of $\hat{\theta}^2$, appropriately centered and scaled (feel free to use heuristic arguments).

Problem 1 Solutions

(a) This model is a subfamily of the bivariate Gaussian location family with no restrictions on μ. That family is full-rank with complete sufficient statistic $T(X) = \sum_i X_i$ and natural parameter μ.

It is clear that

$$\text{Span}\left[\{\mu(\theta_1) - \mu(\theta_2) : \ \theta_1, \theta_2 \in \mathbb{R}\}\right] = \mathbb{R}^2,$$

for example take $\mu(1) - \mu(0) = \binom{1}{1}$ and $\mu(-1) - \mu(0) = \binom{-1}{1}$; as a result $\sum_i X_i$ is minimal sufficient.

Furthermore, taking $f(t) = (t_1/n)^2 - (t_2/n) - n^{-1}$, we have

$$\begin{aligned}\mathbb{E}f(T(X)) &= \mathbb{E}\left[\left(n^{-1}\sum_i X_{i,1}\right)^2\right] - \mathbb{E}[n^{-1}\sum_i X_{i,2}] - n^{-1} \\ &= \theta^2 + n^{-1} - \theta^2 - n^{-1} = 0\end{aligned}$$

But $f(T(X))$ is clearly not almost surely zero, so $T(X)$ is not complete.

(b) The part of the log-likelihood that depends on θ is

$$\begin{aligned}\ell(\theta; X) &= \theta\sum_i X_{i,1} + \theta^2\sum_i X_{i,2} - \theta^2/2 - \theta^4/2 \\ \frac{d\ell}{d\theta}(\theta; X) &= \sum_i X_{i,1} + 2\theta\sum_i X_{i,2} - \theta - 2\theta^3.\end{aligned}$$

The variance of $\sum_i X_{i,1} + 2\theta \sum_i X_{i,2}$ is

$$J_n(\theta) = n + 4\theta^2 n = n(1 + 4\theta^2).$$

(c) The one-sided score test rejects for large values of

$$\frac{\frac{d\ell}{d\theta}(\theta_0)}{\sqrt{J_n(\theta_0)}} = \frac{\sum_i X_{i,1} + 2\theta_0\sum_i X_{i,2} - (\theta_0 + 2\theta_0^3)}{\sqrt{n(1 + 4\theta_0^2)}}.$$

In generic "smooth" models this statistic has mean 0 and variance 1, and converges to $N(0, 1)$, under the null. In this particular model it is exactly $N(0, 1)$ in finite samples, because the $X_{i,j}$ values are independent Gaussian random variables. We reject if it is larger than z_α, which has probability exactly α.

(d) Applying standard results we have

$$\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N\left(0, \frac{1}{1 + 4\theta^2}\right).$$

Further applying the delta method for the function $\theta \mapsto \theta^2$, whose derivative is 2θ, we obtain

$$\sqrt{n}(\hat{\theta}^2 - \theta^2) \Rightarrow N\left(0, \frac{4\theta^2}{1 + 4\theta^2}\right) = N\left(0, \frac{1}{1 + 1/4\theta^2}\right)$$

By contrast, we have $\sqrt{n}\left(\frac{1}{n}\sum_i X_{i,2} - \theta^2\right) \sim N(0,1)$ in finite samples as well as asymptotically, which is a higher variance no matter what θ is. The asymptotic relative efficiency of the "obvious" estimator is therefore $\frac{1}{1+1/4\theta^2}$. Note this is close to 1 as $\theta \to \pm\infty$ but as $\theta \to 0$ it diverges to ∞.

(e) Let $\theta = 0$ and let $n \to \infty$. As we zoom into the origin, eventually the model is nearly a horizontal line through the origin. Because maximum likelihood estimation in a Gaussian location model is equivalent to Euclidean projection, we have $\hat{\theta} \approx \bar{X}_1 = \frac{1}{n}\sum_i X_{i,1}$ and $\hat{\theta}^2 \approx (\bar{X}_1)^2 = \frac{1}{n}\chi_1^2$. If we did all this more carefully we would get

$$n(\hat{\theta}^2 - \theta^2) \Rightarrow \chi_1^2.$$

2. Species abundance (20 points, 5 points / part). Some useful facts for this problem:

- Recall that the Poisson distribution $X \sim \text{Pois}(\lambda)$ has probability mass function
$$\frac{\lambda^x e^{-\lambda}}{x!},$$
on $x = 0, 1, 2, \ldots$. X has mean λ and variance λ.

- The multinomial distribution $(X_1, \ldots, X_d) \sim \text{Multinom}(n, \pi)$ has probability mass function
$$\frac{n!}{\prod_i x_i!} \prod_i \pi_i^{x_i}$$

- Suppose $X_i \overset{\text{ind.}}{\sim} \text{Pois}(\lambda_i)$ for $i = 1, \ldots, d$, and let $X_+ = \sum_i X_i$ and $\lambda_+ = \sum_i \lambda_i$. Then conditional on $X_+ = x_+$, we have
$$(X_1, \ldots, X_d) \sim \text{Multinomial}(x_+, (\lambda_1, \ldots, \lambda_d)/\lambda_+).$$

Consider an ecological sampling problem where we visit m sites and for each of s species, we count the total number of individuals at each site. Let $N_j^{(i)}$ denote the number of individuals of species j at site i. Hence we observe a table of counts of the form

	Species		
Sites	1	$\cdots$	s
1	$N_1^{(1)}$	$\cdots$	$N_s^{(1)}$
$\vdots$	$\vdots$	$N_j^{(i)}$	$\vdots$
m	$N_1^{(m)}$	$\cdots$	$N_s^{(m)}$

We will assume throughout that the rows $N^{(1)}, \ldots, N^{(m)}$ are i.i.d. random vectors in $\mathbb{R}^s$ (but the coordinates $N_1^{(i)}, \ldots, N_s^{(i)}$ within a single site are *not* i.i.d.).

(a) First assume $N_j^{(i)} \overset{\text{ind.}}{\sim} \text{Pois}(\lambda_j)$ for $j = 1, \ldots, s$, and that at each site the species counts are independent, i.e.
$$N^{(i)} \overset{\text{i.i.d.}}{\sim} p_\lambda(n) = \prod_{j=1}^{s} \frac{\lambda_j^{n_j} e^{-\lambda_j}}{n_j!} \tag{1}$$

Find a complete sufficient statistic for the entire data table and give a UMVU estimator for λ_j, the average abundance of species j, explaining why it is UMVU.

(b) Next, assume we use outside data to compute a dissimilarity measure $d(j,k) \in [0, \infty)$ for each pair of species $1 \leq j < k \leq s$; for example $d(j,k)$ could denote how long ago the species diverged in their evolution. Take the $d(j,k)$ as fixed and known.

We might expect that some latent characteristics of the habitat at site i cause similar species to be more or less common together, and we can test this hypothesis by modifying our model:

$$N^{(i)} \overset{\text{i.i.d.}}{\sim} p_{\lambda,\beta}(n) \quad \propto \quad \prod_{j=1}^{s} \frac{\lambda_j^{n_j} e^{-\lambda_j}}{n_j!} \quad \times \prod_{1 \leq j < k \leq s} \exp\{\beta e^{-d(j,k)} n_j n_k\}. \quad (2)$$

Show that this model is an exponential family with $s+1$ sufficient statistics, and find the natural parameter corresponding to each (as always, there are multiple ways to write these). You do *not* need to find the normalizing constant.

(c) Find a UMPU test of $H_0: \ \beta = 0$ (independence) vs. $H_1: \ \beta > 0$ (positive correlation between similar species) and explain how to find its critical value.

(d) (*) Now suppose we want to make our test more robust by dropping the Poisson assumption: under the null hypothesis the species counts are still independent, but now with unknown distributions (still supported on the non-negative integers):

$$N^{(i)} \overset{\text{i.i.d.}}{\sim} \prod_{j=1}^{s} F_j(n_j) \quad (\text{under } H_0),$$

and under the alternative the counts of similar species are still more correlated. Modify your test from part (b) so that it controls finite-sample Type I error, in this nonparametric model.

Problem 2 solutions:

1. Let $U_j = \sum_i N_j^{(i)}$. The full model density is

$$\prod_{i=1}^{m}\prod_{j=1}^{s} \frac{\lambda_j^{n_j^{(i)}} e^{-\lambda_j}}{n_j^{(i)}!} = \exp\left\{\sum_j \eta_j U_j - m\sum_j e^{\eta_j}\right\} \frac{1}{\prod_{i,j} n_j^{(i)}!},$$

where $\eta_j = \log \lambda_j$ and $U_j = \sum_i N_j^{(i)}$. This is a full-rank exponential family with η_j ranging over all of $\mathbb{R}^s$. The s sufficient statistics $U_1, \ldots, U_s$ are therefore complete sufficient. As U_j/m is unbiased for λ_j, it is UMVU.

2. Let $T = \sum_i \sum_{1\le j<k\le s} e^{-d(j,k)} N_j^{(i)} N_k^{(i)}$. The full model density is now proportional to

$$\exp\left\{\beta T + \sum_j \eta_j U_j\right\} \frac{1}{\prod_{i,j} n_j^{(i)}!},$$

which has $s+1$ sufficient statistics $U_1, \ldots, U_s, T$ with corresponding natural parameters $\eta_1, \ldots, \eta_s, \beta$.

3. As usual, the UMPU test will reject for large values of T conditional on U. In practice, we can find the critical value by Monte Carlo, by resampling the entire data set under the null model, conditional on U_j, and recomputing the test statistic T each time we do. Due to the fact about conditional multinomials, and the fact that species are independent of one another under the null, this means that for each species (each column of the table) we should resample $(N_1^{(j)}, \ldots, N_m^{(j)}) \overset{\text{ind.}}{\sim} \text{Multinomial}(U_j, (1, \ldots, 1)/m)$, for $j = 1, \ldots, m$.

   One way to look at this is that we take every individual that was counted in the data set, and we independently send each of them to a uniformly random site (while preserving which species they are). Note this reassignment does not change the total number of individuals of each species in the data set.

4. In this case the null model is essentially that $(N_j^{(1)}, \ldots, N_j^{(s)}) \overset{\text{i.i.d.}}{\sim} F_j$, independently for each j with F_j completely unknown. The complete sufficient statistic for the (null) model is therefore the empirical distribution of counts $(N_j^{(1)}, \ldots, N_j^{(s)})$ for each of the s species (i.e., s

empirical distributions each recording m observations). Resampling the data table conditional on those empirical distributions amounts to randomly shuffling the entries of each column of the table and then recompouting the test statistic. This is a permutation test where we permute each column of the table separately.

3. Inverse gamma prior (20 points, 4 points / part). Some useful facts for this problem:

- Recall that the Gaussian density function for $Z \sim N(\mu, \sigma^2)$ is

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- A χ^2_d random variable has mean d and variance $2d$.

- If Y is a Gamma(α, β) random variable (in its "rate parameterization") then it has density

$$\frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\},$$

  on $(0, \infty)$. Y has mean α/β and variance α/β^2. This distribution is defined for $\alpha, \beta > 0$.

- The inverse-gamma distribution (denoted $IG(\alpha, \beta)$) is the distribution of $W = 1/Y$ where $Y \sim$ Gamma(α, β). Then $W \in (0, \infty)$ has the density

$$\frac{\beta^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} \exp\{-\beta/w\}.$$

  Note that β is a scale parameter for W. W has mean $\frac{\beta}{\alpha-1}$ provided $\alpha > 1$, and variance $\frac{\beta-1}{(\alpha-1)^2(\alpha-2)}$ provided $\alpha > 2$. This distribution is likewise defined for $\alpha, \beta > 0$.

- Define the *squared relative error* loss function

$$L_{\text{rel}}(d, \theta) = \left(\frac{d-\theta}{\theta}\right)^2 = \left(\frac{d}{\theta} - 1\right)^2,$$

  and define the corresponding risk function $R_{\text{rel}}(\delta(\cdot), \theta) = \mathbb{E}_\theta[L_{\text{rel}}(\delta(X), \theta)]$.

  Consider the Bayesian model with

$$\theta \sim IG(\alpha, \beta),$$
$$X_1, \ldots, X_n \mid \theta \overset{\text{i.i.d.}}{\sim} N(0, \theta)$$

Note that the variance is θ, not θ^2, and assume $n \geq 2$.

(a) Find the posterior distribution of θ given $X = (X_1, \ldots, X_n)$ and the Bayes estimator for θ under the usual squared error loss.

(b) Give the mean squared error of the Bayes estimator from part (a), as a function of θ (you don't need to try too hard to simplify it).

(c) Find the Bayes estimator for θ under the squared relative error loss L_{rel}.

(d) (*) For the estimator in part (c), find the risk function $R_{\text{rel}}(\delta(\cdot), \theta)$ as a function of θ and show that the Bayes risk is $\frac{2}{n+2(\alpha+1)}$.

(e) For the relative squared error risk, find a linear estimator of the form $\delta(X) = a \sum_{i=1}^{n} X_i^2$ that is minimax, and prove it is minimax.

Problem 3 solutions:

(a) Let $T = \sum_i X_i^2$, which is the complete sufficient statistic for the Gaussian scale family.

Multiplying the prior by the likelihood and dropping factors that do not depend on θ, we have

$$\begin{aligned}\theta \mid X &\propto \theta^{-\alpha-1} e^{-\beta/\theta} \cdot \frac{1}{\theta^{n/2}} e^{-T/2\theta} \\ &= \theta^{-\alpha-n/2-1} \exp\{-(\beta + T/2)/\theta\} \\ &\propto IG(\alpha + n/2, \beta + T/2) \\ &= IG(\tilde{\alpha}, \tilde{\beta})\end{aligned}$$

where $\tilde{\alpha} = \alpha + n/2$ and $\tilde{\beta} = \beta + T/2$.

The Bayes estimator for squared error loss is just the posterior mean:

$$\delta_1(X) = \mathbb{E}[\theta \mid X] = \frac{\tilde{\beta}}{\tilde{\alpha} - 1} = \frac{2\beta + T}{2(\alpha - 1) + n}$$

(b) We will use the fact that $T = \sum_i X_i^2 \sim \theta\chi_n^2$, which has mean $n\theta$ and variance $2n\theta^2$. The bias is

$$\begin{aligned}\text{Bias}_\theta &= \mathbb{E}_\theta \delta_1(X) - \theta \\ &= \frac{2\beta + n\theta}{2(\alpha - 1) + n} - \theta \\ &= \frac{2\beta - 2(\alpha - 1)\theta}{2(\alpha - 1) + n}.\end{aligned}$$

The variance is

$$\begin{aligned}\text{Var}_\theta &= \frac{2n\theta^2}{(2(\alpha - 1) + n)^2}, \quad \text{and} \\ \text{MSE}_\theta &= \text{Bias}_\theta^2 + \text{Var}_\theta \\ &= \frac{4(\beta - (\alpha - 1)\theta)^2 + 2n\theta^2}{(2(\alpha - 1) + n)^2}\end{aligned}$$

(c) We want to solve

$$\min_d \mathbb{E}\left[\left(\frac{d - \theta}{\theta}\right)^2 \mid X\right],$$

over $d \in (0, \infty)$. Expanding the square, we get

$$\min_d d^2 \mathbb{E}[\theta^{-2} \mid X] - 2d\mathbb{E}[\theta^{-1} \mid X] + 1,$$

leading to the solution

$$d^* = \frac{\mathbb{E}[\theta^{-1} \mid X]}{\mathbb{E}[\theta^{-2} \mid X]}.$$

Now using the fact that $\theta^{-1} \sim \text{Gamma}(\alpha, \beta)$, we have

$$\begin{aligned}
\mathbb{E}[\theta^{-1} \mid X] &= \tilde{\alpha}/\tilde{\beta} \\
\mathbb{E}[\theta^{-2} \mid X] &= \text{Var}[\theta^{-1} \mid X] + \mathbb{E}[\theta^{-1} \mid X]^2 \\
&= \tilde{\alpha}/\tilde{\beta}^2 + (\tilde{\alpha}/\tilde{\beta})^2. \\
\delta_2(X) &= \frac{\tilde{\alpha}/\tilde{\beta}}{\tilde{\alpha}/\tilde{\beta}^2 + (\tilde{\alpha}/\tilde{\beta})^2} \\
&= \frac{\tilde{\beta}}{1 + \tilde{\alpha}} \\
&= \frac{2 + T}{2(\alpha + 1) + n}.
\end{aligned}$$

(d) We can essentially repeat the same computation from part (b), replacing $\alpha - 1$ with $\alpha + 1$, to obtain

$$\text{MSE}_\theta(\delta_2) = \frac{4(\beta - (\alpha + 1)\theta)^2 + 2n\theta^2}{(2(\alpha + 1) + n)^2}.$$

Furthermore,

$$\begin{aligned}
R_{\text{rel}}(\theta) &= \mathbb{E}_\theta \left[\left(\frac{d - \theta}{\theta} \right)^2 \right] \\
&= \frac{\text{MSE}_\theta(\delta_2)}{\theta^2} \\
&= \frac{4(\beta/\theta - (\alpha + 1))^2 + 2n}{(2(\alpha + 1) + n)^2}.
\end{aligned}$$

Now integrating over θ to obtain the Bayes risk, we get

$$\mathbb{E}_{\alpha,\beta}[R_{\text{rel}}(\theta)] = \frac{2n + 4\mathbb{E}_{\alpha,\beta}\left[(\beta/\theta - (\alpha + 1))^2\right]}{(2(\alpha + 1) + n)^2}.$$

Now we can use the fact that

$$\mathbb{E}_{\alpha,\beta}[\beta/\theta] = \mathrm{Var}_{\alpha,\beta}[\beta/\theta] = \alpha,$$

to obtain

$$\mathbb{E}_{\alpha,\beta}\left[(\beta/\theta - (\alpha+1))^2\right] = \alpha + 1.$$

Plugging into the previous expression, we get

$$\mathbb{E}_{\alpha,\beta}[R_{\mathrm{rel}}(\theta)] = \frac{2n + 4(\alpha+1)}{(2(\alpha+1)+n)^2} = \frac{2}{n + 2(\alpha+1)}$$

(e) The relative squared error risk for a linear estimator is

$$\begin{aligned}
\mathbb{E}_\theta\left[\left(\frac{a\sum_i X_i^2 - \theta}{\theta}\right)^2\right] &= \mathbb{E}_\theta\left[\left(\frac{a}{\theta}\sum_i X_i^2 - 1\right)^2\right] \\
&= (an-1)^2 + 2a^2 n \\
&= a^2 n(n+2) - 2an + 1
\end{aligned}$$

using the fact that $\frac{a}{\theta}\sum_i X_i^2$ has mean an and variance $2a^2n$. Minimizing the risk gives $a^* = \frac{1}{n+2}$, so the best linear estimator (for relative risk) is $\frac{1}{n+2}\sum_i X_i^2$, which has risk $1 - \frac{n}{n+2} = \frac{2}{n+2}$. Moreover, this risk is constant in θ.

Because the Bayes risk of any Bayes estimator gives a lower bound on the minimax risk, we know that the minimax risk is at least $\sup_{\alpha>0} \frac{2}{n+2(\alpha+1)} = \frac{2}{n+2}$. But our optimal linear estimator achieves constant risk of $\frac{2}{n+2}$, giving a matching upper bound; hence our estimator $\frac{1}{n+2}\sum_i X_i^2$ is indeed achieving the minimax risk.

Note: there was a lot of confusion about “constant risk” on this problem. The Bayes risk (average-case risk) is always a constant; we get it by averaging over θ so there’s no way it could possibly depend on θ. Similarly, any a we choose will give us a constant relative risk function (try it); but some of those constant risk functions are strictly better than others, so they can’t all be minimax! What we know is: if a *Bayes estimator* has a constant *risk function* in θ, that’s when we know we have a minimax estimator. That doesn’t actually apply in this problem because the risk functions of the Bayes estimators are not constant, and our linear estimators (which do have constant risk functions) aren’t among the Bayes estimators in the problem.

4. Inference in the Laplace family (15 points, 5 points / part). Some useful facts for this problem:

- The Laplace location family with location parameter θ is given by the density $p_\theta(x) = f(x-\theta)$, where $f(x) = \frac{1}{2}e^{-|x|}$.
- The sign function is defined as

$$\operatorname{sign}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}.$$

Assume we observe $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Laplace}(\theta)$.

(a) Find the score test of $H_0: \theta = 0$ vs. $H_1: \theta > 0$. Give the test statistic and threshold value in terms of a quantile of a binomial distribution (for simplicity you may assume α is chosen so the binomial distribution has an exact α quantile, so the test need not be randomized. Also note $\mathbb{P}_\theta(X = 0) = 0$ for all θ, so you don't need to worry about what happens there).

(b) Suppose we are not so sure about the Laplace assumption, but we do believe the data come from a symmetric location family, meaning $p_\theta(x) = f(x-\theta)$ for some unknown $f : \mathbb{R} \to [0, \infty)$ that integrates to 1 and is symmetric about the origin (we can think of the nonparametric family as being parameterized by (θ, f)). Show that the test from part (a) is still a valid, finite-sample test of $H_0: \theta = 0$ vs. $H_1: \theta > 0$ in this larger family.

(c) Now consider testing $H_0: \theta \leq 0$ vs. $H_1: \theta > 0$ (note the null hypothesis now includes negative values of θ). Show that the test from part (a) is a valid and unbiased level-α test for the nonparametric family from part (b).

Problem 4 solutions:

(a) The part of the log-likelihood that depends on θ is

$$\ell(\theta) = -\sum_i |X_i - \theta|$$
$$\frac{d\ell}{d\theta}(\theta) = \text{sign}(X_i - \theta).$$

Evaluating at $\theta = 0$, we get that the score is $S(X) = \sum_i \text{sign}(X_i)$. Slightly modifying the statistic in a way that doesn't change the test, we can take

$$T(X) = (S(X) + n)/2 = \sum_i 1\{X_i > 0\} \sim \text{Binom}(n, 1/2).$$

We reject if $T(X)$ is above the upper-α quantile of this binomial distribution.

(b) Generically, we have

$$T(X) \sim \text{Binom}(n, \pi_{theta,f})$$

where

$$\pi_{\theta,f} = \mathbb{P}_{X_1 \overset{\text{i.i.d.}}{\sim} f(x-\theta)}(X_1 > 0) = \int_{-\theta}^{\infty} f(u)\, du.$$

For any symmetric f, it is clear that $\pi_{0,f} = 1/2$, so the null distribution of $T(X)$ does not depend on f and the test has exact level α as in part (a).

(c) The integral $\pi_{\theta,f}$ is evidently increasing in θ for any function f with $\pi_{0,f} = 1/2$, and the binomial random variable $T(X)$ is stochastically increasing in $\pi_{\theta,f}$, so the rejection probability is increasing as well, and equals exactly α and $\theta = 0$ for any f. As a result, the test rejects with probability no greater than α if $\theta \leq 0$ and with probability no less than α if $\theta > 0$.