

UNIVERSITETET I OSLO

Matematisk Institutt

EXAM IN: **STK 4011/9011 – Statistical Inference Theory**
WITH: **Nils Lid Hjort**
AUXILIA: **One single sheet of paper with the candidate's
own personal hand-written notes
Calculator**
TIME FOR EXAM: **Wednesday 18/xii/2024, 9:00–13:00**

This exam set contains four exercises and comprises four pages. Write your solutions in bokmål, nynorsk, riksmål, Danish, Swedish, English, or Latin.

Exercise 1: binomials and the arcsine transformation

THE ARCSINE IS ASININE, claims a 2010 article in the journal *Ecology*. Below we deal with approximations to the binomial distribution, indeed also via the arcsine function.

- (a) Suppose Y is binomial (n, p) , which we write as $Y \sim \text{binom}(n, p)$, with n the number of experiments carried out and $p \in (0, 1)$ the probability of the event in question. With $\hat{p} = Y/n$ the traditional estimator, explain via the Central Limit Theorem that $\sqrt{n}(\hat{p} - p) \rightarrow_d X$, where $X \sim N(0, p(1 - p))$. If one observes $Y = 33$ for $n = 100$, explain how you can use this approximate normality to set an approximate 95 percent confidence interval for p .
- (b) The function $y = \arcsin x$ is the inverse of the sinus function, i.e. the solution to $x = \sin y$. Its derivative is $1/(1 - x^2)^{1/2}$, for $x \in (-1, 1)$. Show that with $\gamma = 2 \arcsin(p^{1/2})$ and $\hat{\gamma} = 2 \arcsin(\hat{p}^{1/2})$, we have

$$Z_n = \sqrt{n}(\hat{\gamma} - \gamma) \rightarrow_d N(0, 1).$$

Explain how you can use this to construct another approximate 95 percent confidence interval for p .

- (c) Suppose there are six different independent binomial experiments, leading to $Y_j \sim \text{binom}(n, p_j)$ for $j = 1, \dots, 6$, for simplicity assuming the n to be the same, giving also rise to estimators $\hat{p}_j = Y_j/n$. Show that with $\gamma_j = 2 \arcsin(p_j^{1/2})$ and $\hat{\gamma}_j = 2 \arcsin(\hat{p}_j^{1/2})$, we have

$$\begin{pmatrix} X_{n,1} \\ \vdots \\ X_{n,6} \end{pmatrix} = \begin{pmatrix} \sqrt{n}(\hat{\gamma}_1 - \gamma_1) \\ \vdots \\ \sqrt{n}(\hat{\gamma}_6 - \gamma_6) \end{pmatrix} \rightarrow_d \begin{pmatrix} X_1 \\ \vdots \\ X_6 \end{pmatrix},$$

where $X_1, \dots, X_6$ are independent standard normals.

(d) Under the null hypothesis H_0 that $p_1 = \dots = p_6$, show that

$$Q_n = \sum_{j=1}^6 n(\hat{\gamma}_j - \gamma^*)^2 \rightarrow_d \chi_5^2,$$

where γ^* is the average $(1/6) \sum_{j=1}^6 \hat{\gamma}_j$. Explain how this can be used to construct a 5 percent level test for H_0 .

Exercise 2: profiling with confidence

PEOPLE ARE SOMETIMES PROFILED BASED ON THEIR PROFILES. Likelihood profiling is useful, though. Consider the classic setup with $Y_1, \dots, Y_n$ being i.i.d. from the normal $N(\xi, \sigma^2)$ distribution. The usual estimators are the sample mean and standard deviation, i.e. $\hat{\xi} = \bar{Y}$ and $\hat{\sigma} = \{Q_0/(n-1)\}^{1/2}$, where we write $Q_0 = \sum_{i=1}^n (Y_i - \bar{Y})^2$.

(a) We know from the core curriculum that the variable $t = \sqrt{n}(\hat{\xi} - \xi)/\hat{\sigma}$ has the t distribution with $m = n - 1$ degrees of freedom. With G_m the cumulative distribution function for this t_m , explain that

$$C_n(\xi, \text{data}) = G_m(\sqrt{n}(\xi - \hat{\xi}_{\text{obs}})/\hat{\sigma}_{\text{obs}})$$

is an exact confidence distribution, in which $\hat{\xi}_{\text{obs}}$ and $\hat{\sigma}_{\text{obs}}$ are the observed values of these estimators.

(b) Show that the log-likelihood function may be written

$$\ell_n(\xi, \sigma) = -n \log \sigma - \frac{1}{2} \{Q_0 + n(\xi - \hat{\xi})^2\} / \sigma^2 - \frac{1}{2} n \log(2\pi).$$

Find the maximum likelihood (ML) estimators $\hat{\xi}_{\text{ml}}$ and $\hat{\sigma}_{\text{ml}}$, and also an expression for the maximised log-likelihood, $\ell_{n,\text{max}}$.

(c) Then work out an expression for the log-likelihood profile

$$\ell_{n,\text{prof}}(\xi) = \max\{\ell_n(\xi, \sigma) : \text{all } \sigma\}.$$

Show that this leads to the deviance function

$$D_n(\xi) = 2\{\ell_{n,\text{max}} - \ell_{n,\text{prof}}(\xi)\} = n \log \left(1 + \frac{1}{n-1} \frac{n(\xi - \hat{\xi})^2}{\hat{\sigma}^2} \right).$$

(d) Explain from the curriculum that the Wilks theorem implies that with $\xi = \xi_0$ the correct value of ξ , the $D_n(\xi_0)$ tends to a chi-squared in distribution. What is the degrees of freedom, for this limiting chi-squared? Now verify this limit distribution statement, directly, via the formula above. You may use the approximation $\log(1 + a) \doteq a - \frac{1}{2}a^2$, valid for small a .

- (e) We started out in (a) with an exact confidence distribution for ξ , via the t_m distribution. Explain nevertheless how the Wilks theorem may be used here, to construct an approximate confidence distribution, $cc_{n,w}(\xi, \text{data})$.

Exercise 3: tilting distributions

A CERTAIN PHENOMENON DELIVERS NONNEGATIVE OUTCOMES following the cumulative distribution function $F_0(y) = y/(1+y)$ for $y \geq 0$, but sometimes they rather follow the tilted distribution

$$P(Y \leq y) = F(y, \theta) = \left(\frac{y}{y+1} \right)^\theta \quad \text{for } y \geq 0,$$

with θ a positive parameter, perhaps not far from one.

- (a) Show that the density may be written

$$f(y, \theta) = \theta \frac{y^{\theta-1}}{(1+y)^{\theta+1}} \quad \text{for } y \geq 0.$$

Find the score function $u(y, \theta)$, and show that the Fisher information function becomes $J(\theta) = 1/\theta^2$.

- (b) Suppose $Y_1, \dots, Y_n$ are i.i.d. from this model. Give the Cramér–Rao lower bound for unbiased estimators of θ .
- (c) Find the log-likelihood function $\ell_n(\theta)$, and show that the maximum likelihood estimator can be expressed as $\hat{\theta}_{\text{ml}} = 1/W_n$, where $W_n = n^{-1} \sum_{i=1}^n \log(1 + 1/Y_i)$. Identify the limit distribution of $\sqrt{n}(\hat{\theta}_{\text{ml}} - \theta)$, and comment on your answer.
- (d) One wishes to test the hypothesis $H_0: \theta = 1$ versus the alternative that $\theta > 1$, with significance level say 0.05. Show that the power optimal test takes the form of rejecting H_0 when W_n is sufficiently small, say $W_n \leq c_n$. What is the formal requirement for the rejection threshold c_n ?
- (e) To find the c_n , show first that $V_i = \log(1 + 1/Y_i)$ has the exponential density $\theta \exp(-\theta v)$. Using the fact, from this, that $2\theta V_i \sim \chi_2^2$, show that W_n can be represented as $(1/\theta)\chi_{2n}^2/(2n)$.
- (f) From this again, show that $2nc_n = \gamma_{0.05, 2n} = \Gamma_{2n}^{-1}(0.05)$, the 0.05 quantile in the χ_{2n}^2 distribution, using $\Gamma_{2n}(\cdot)$ for the cumulative distribution function. Find an expression for the power function of the test.
- (g) Find the maximum likelihood estimator for the median $F^{-1}(\frac{1}{2}, \theta)$.
- (h) For the final point in this exercise, introduce a Bayesian element, with θ having the Gamma distribution prior with some prior parameters (a, b) , with density proportional to $\theta^{a-1} \exp(-b\theta)$. It is known that the mean and variance of this distribution are a/b and a/b^2 . Your task is to find a formula for the Bayes estimator $\hat{\theta}_B = E(\theta | \text{data})$.

Exercise 4: progressive regression

IT IS THE MARK OF A PRIMITIVE SOCIETY TO VIEW REGRESSION AS PROGRESS (claims B.D. Walsch). Statistical progress needs good regression models and methods, however. Suppose that independent regression pairs (x_i, Y_i) are available, for $i = 1, \dots, n$, where Y_i given x_i has the distribution studied in Exercise 3, but with θ_i potentially influenced by x_i , in this fashion:

$$P(Y_i \leq y) = \left(\frac{y}{y+1} \right)^{\theta_i} \quad \text{with } \theta_i = a + b(x_i - \bar{x}) \text{ for } i = 1, \dots, n.$$

Here the x_i are seen as given covariates; we write $\bar{x} = (1/n) \sum_{i=1}^n x_i$; and a and b are unknown parameters.

- (a) Explain briefly how the parameters a and b can be interpreted here. For given a positive, what is the parameter range for b ?
- (b) Write down an expression for the log-likelihood function, say $\ell_n(a, b)$. Identify a sufficient set of statistics.
- (c) Then explain, perhaps in a couple of paragraphs, how you would proceed, analysing a concrete dataset with such (x_i, Y_i) pairs. In particular, outline how you could test the hypothesis $H_0: b = 0$ versus $b \neq 0$.