

## Homework 6

ALECK ZHAO

October 21, 2016

1. Prove the converse of the factorization theorem, namely prove that if  $T$  is a sufficient statistic, then the joint density can be factored as

$$f(x_1, \dots, x_n | \theta) = g(T, \theta)h(x_1, \dots, x_n)$$

Also show that if  $T$  is sufficient for  $\theta$ , then the MLE must be a function of  $T$ .

2. Let  $\hat{\theta}$  be an estimator for a parameter  $\theta$ , and suppose that  $\text{Var}(\hat{\theta}) < \infty$ . Let  $T$  be a sufficient statistic for  $\theta$ . Consider the random variable

$$Y = E[\hat{\theta} | T]$$

Prove that

$$E[(Y - \theta)^2] \leq E[(\hat{\theta} - \theta)^2]$$

3. Complete all the details of the example we discussed in lecture. Let  $X_1, \dots, X_n$  be iid data from a normal distribution with unknown mean and known variance  $\sigma^2$ . Suppose that  $\theta$  is assumed to be random, with prior distribution also normal; assume that the mean and variance of the prior distribution of  $\theta_0$  and  $\sigma_{pr}^2$ , where both  $\sigma_0$  and  $\sigma_{pr}^2$  are known.

- (a) Compute the posterior distribution

$$f_{\theta|\mathbf{X}}(\theta | x_1, \dots, x_n)$$

where  $\mathbf{X} = (X_1, \dots, X_n)$ , and specify all the parameters of this distribution.

- (b) For what value of  $\theta$  is this posterior density maximized? Given this, what would you choose as an estimate for  $\theta$ ?
  - (c) How do the prior variance  $\sigma_{pr}^2$  and the posterior variance compare? Which one is larger? Does this make sense? Why?
  - (d) How does the estimator you obtained in part b compare to the MLE?
4. Suppose we are in the Bayesian framework and we wish to estimate a parameter  $\theta$  with prior distribution  $f$  from some family of distributions  $G$ . If, conditional on the value of the parameter, the data have some distribution  $H$  and the posterior distribution is again in the family  $G$ , we say that  $G$  and  $H$  are conjugate.

- (a) Show that if  $X_i$  are iid Bernoulli ( $p$ ) and  $p$  has a Beta-distributed prior, so that

$$f_p(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

where, as usual,

$$\Gamma(u) = \int_0^\infty t^{u-1} e^{-t} dt$$

then the Bernoulli and Beta families are conjugate.

- (b) What if the  $X_i$  are binomial with parameters  $n, p$  where  $n$  is known and  $p$  has, again, a Beta distribution? Are the binomial and Beta families conjugate?
- (c) Show that if  $X_i$  are iid exponential with parameter  $\lambda$ , and  $\lambda$  has a Gamma-distributed prior, then the posterior also has a Gamma distribution. What is a reasonable estimate for  $\lambda$  in this Bayesian setting? How does it compare to the MLE for the exponential?
5. Suppose we observe an iid sample  $X_1, \dots, X_n$  from the distribution that is uniform in the interval  $[-\theta, \theta]$  for some unknown  $\theta > 0$ .
- (a) Find the MLE for  $\theta$ .
- (b) Show that the pair  $T = \max\{X_1, \dots, X_n\}$  and  $S = \min\{X_1, \dots, X_n\}$  are sufficient for  $\theta$ .
6. Suppose  $(U, V)$  is a uniformly distributed point in the unit circle  $\{(x, y) \mid x^2 + y^2 \leq 1\}$  in the plane.
- (a) Determine the marginal PDFs of  $U$  and  $V$  and expectations  $E[U]$  and  $E[V]$ . Also determine the covariance  $\text{Cov}(U, V)$  and decide if  $U, V$  are independent.

*Solution.* The area of the unit circle is  $\pi$ , so the joint density is given by

$$f_{U,V}(u, v) = \frac{1}{\pi}$$

The marginal density of  $u$  is given by

$$f_U(u) = \int f_{U,V}(u, v) dv = \int_{-\sqrt{1-u^2}}^{\sqrt{1-u^2}} \frac{1}{\pi} dv = \frac{2\sqrt{1-u^2}}{\pi}$$

Similarly, the marginal density of  $v$  is given by

$$f_V(v) = \frac{2\sqrt{1-v^2}}{\pi}.$$

It's easy to see that these densities are symmetric about the origin, so  $E[U] = E[V] = 0$ . The covariance is given by

$$\begin{aligned} \text{Cov}(U, V) &= E[UV] - E[U]E[V] = E[UV] \\ &= \int \int uv \cdot f_{U,V}(u, v) dv du \\ &= \frac{1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-u^2}}^{\sqrt{1-u^2}} uv dv du \\ &= 0 \end{aligned}$$

but the product of the marginal densities is

$$f_U(u)f_V(v) = \frac{2\sqrt{1-u^2}}{\pi} \cdot \frac{2\sqrt{1-v^2}}{\pi} = \frac{4(1-u^2)(1-v^2)}{\pi^2} \neq f_{U,V}(u, v)$$

so  $U$  and  $V$  are not independent.

□

- (b) Let  $W = U^2 + V^2$ . Compute the density  $f_W(w)$  for  $W$ .

*Solution.* Consider the probability  $F_W(w) = P(W \leq w) = P(U^2 + V^2 \leq w)$ . This is a circle of radius  $w$  centered at the origin, but  $U^2 + V^2$  can be anywhere in the unit circle, so this probability is given by

$$P(W \leq w) = \frac{w^2 \pi}{\pi} = w^2$$

so the density is given by

$$f_W(w) = \frac{d}{dw} F_W(w) = \frac{d}{dw} [w^2] = 2w, \quad 0 \leq 1 \leq w$$

□

- (c) Let  $R = \theta U$ , and  $T = \theta V$ , where  $\theta > 0$  is some non-random parameter. Compute the joint distribution of  $(R, T)$ .

*Solution.* We have

$$f_{R,T}(r, t) = f_{U,V}(u, v) \left| \frac{d(u, v)}{d(r, t)} \right|$$

where  $U = R/\theta$  and  $V = T/\theta$ , so the joint density of  $R, T$  is given by

$$f_{R,T}(r, t) = \frac{1}{\pi} \left| \begin{bmatrix} 1/\theta & 0 \\ 0 & 1/\theta \end{bmatrix} \right| = \frac{1}{\theta^2 \pi}$$

□

7. Suppose we observe independent pairs  $(X_i, Y_i)$  where each  $(X_i, Y_i)$  has a uniform distribution in the circle of unknown radius  $\theta$  and centered at  $(0, 0)$  in the plane.

- (a) Show that  $(X_i/\theta, Y_i/\theta)$  has a uniform distribution in the unit circle, and find the PDF of  $X_i^2 + Y_i^2$ .

*Proof.* The joint density of  $X_i, Y_i$  is given by

$$f_{X_i, Y_i}(x_i, y_i) = \frac{1}{\theta^2 \pi}$$

so letting  $X_i = \theta A$ ,  $Y_i = \theta B$ , we have the joint density of  $A, B$  is

$$f_{A,B}(a, b) = f_{X_i, Y_i}(x_i, y_i) \left| \frac{d(x_i, y_i)}{d(a, b)} \right| = \frac{1}{\theta^2 \pi} \left| \begin{bmatrix} \theta & 0 \\ 0 & \theta \end{bmatrix} \right| = \frac{1}{\pi}$$

which is exactly the joint density of a uniform distribution on the unit circle, as desired.

Let  $W = X_i^2 + Y_i^2$ . Then the CDF of  $W$  is given by

$$F_W(w) = P(W \leq w) = P(X_i^2 + Y_i^2 \leq w)$$

which is a circle of radius  $w$  centered on the origin, and since  $X_i, Y_i$  is uniformly distributed on a circle of radius  $\theta$ , this probability is

$$F_W(w) = \frac{w^2 \pi}{\theta^2 \pi} = \frac{w^2}{\theta^2}.$$

Thus, the density of  $W$  is given by

$$f_W(w | \theta) = \frac{d}{dw} F_W(w) = \frac{d}{dw} \left[ \frac{w^2}{\theta^2} \right] = \frac{2w}{\theta^2}, \quad 0 \leq w \leq \theta$$

□

- (b) Show that  $(X_1^2 + Y_1^2, \dots, X_n^2 + Y_n^2)$  is a sufficient statistic for  $\theta$ .

*Proof.* Let  $W_i = X_i^2 + Y_i^2$ . Then the likelihood function is given by

$$f(W_1, \dots, W_n | \theta) = \prod_{i=1}^n f(W_i | \theta) = \prod_{i=1}^n \frac{2w_i}{\theta^2} = \frac{2^n}{\theta^{2n}} \prod_{i=1}^n w_i$$

□

i++i

- (c) Find the MLE and determine its density function and its bias. Are the regularity assumptions required on the MLE satisfied here?

*Solution.* As above, the joint density

$$f[(X_1, Y_1), \dots, (X_n, Y_n) | \theta] = \prod_{i=1}^n f[(X_i, Y_i) | \theta] = \prod_{i=1}^n \frac{1}{\theta^2 \pi} = \frac{1}{\theta^{2n} \pi^n}$$

Since  $X_i^2 + Y_i^2 \leq \theta^2$ , the MLE  $\hat{\theta}$  is

$$\hat{\theta} = \max_{1 \leq i \leq n} \sqrt{X_i^2 + Y_i^2}$$

Consider the CDF of  $\hat{\theta}$

$$\begin{aligned} F(t) &= P(\hat{\theta} \leq t) = P\left(\max_{1 \leq i \leq n} \sqrt{X_i^2 + Y_i^2} \leq t\right) \\ &= P\left(\sqrt{X_1^2 + Y_1^2}, \dots, \sqrt{X_n^2 + Y_n^2} \leq t\right) \\ &= \prod_{i=1}^n P\left(\sqrt{X_i^2 + Y_i^2} \leq t\right) \\ &= \prod_{i=1}^n P(W_i \leq t^2) = \prod_{i=1}^n \frac{t^2}{\theta^2} = \frac{t^{2n}}{\theta^{2n}} \end{aligned}$$

and the density of  $\hat{\theta}$  is the derivative of this wrt to  $t$ :

$$f_{\hat{\theta}}(t) = \frac{\partial}{\partial t} \left[ \frac{t^{2n}}{\theta^{2n}} \right] = \frac{2nt^{2n-1}}{\theta^{2n}}$$

Then  $E[\hat{\theta}]$  is given by

$$\begin{aligned} E[\hat{\theta}] &= \int_0^\theta t \frac{2nt^{2n-1}}{\theta^{2n}} dt = \int_0^\theta \frac{2nt^{2n}}{\theta^{2n}} dt \\ &= \frac{2nt^{2n+1}}{\theta^{2n}(2n+1)} \Big|_0^\theta = \frac{2n\theta}{(2n+1)} \end{aligned}$$

so the bias of  $\hat{\theta}$  is

$$E[\hat{\theta}] - \theta = \frac{2n\theta}{2n+1} - \theta = -\frac{\theta}{2n+1}.$$

The support of the distribution of  $(X_i, Y_i)$  is

$$\{(x_i, y_i) | f[(x_i, y_i) | \theta] > 0\} = \{(x_i, y_i) | 1/\theta^2 \pi > 0\}$$

which is the entire domain, and doesn't depend on  $\theta$ . Thus the MLE satisfies the regularity conditions.

□

- (d) Compute the variance of the MLE and simplify it so that it is clear how this variance decays with the sample size  $n$ .

*Solution.* The variance of the MLE is given by

$$\text{Var}(\hat{\theta}) = E[\hat{\theta}^2] - (E[\hat{\theta}])^2$$

where

$$\begin{aligned} E[\hat{\theta}^2] &= \int_0^\theta t^2 \frac{2nt^{2n-1}}{\theta^{2n}} dt = \int_0^\theta \frac{2nt^{2n+1}}{\theta^{2n}} dt \\ &= \frac{2nt^{2n+2}}{\theta^{2n}(2n+2)} \Big|_0^\theta = \frac{n\theta^2}{n+1} \end{aligned}$$

so the variance is

$$\begin{aligned} \text{Var}(\hat{\theta}) &= \frac{n\theta^2}{n+1} - \left( \frac{2n\theta}{2n+1} \right)^2 \\ &= \theta^2 \left( \frac{n}{n+1} - \frac{4n^2}{(2n+1)^2} \right) = \frac{n\theta^2}{(n+1)(2n+1)^2} \end{aligned}$$

Clearly, this diminishes very quickly as  $n$  increases. □

- (e) Find the MSE of the MLE. As  $n \rightarrow \infty$ , which term contributes more to the MSE, the squared bias or the variance?

*Solution.* The MSE is given by

$$\begin{aligned} E[(\hat{\theta} - \theta)^2] &= \text{Var}(\hat{\theta}) + \left( E[\hat{\theta} - \theta] \right)^2 \\ &= \frac{n\theta^2}{(n+1)(2n+1)^2} + \left( -\frac{\theta}{2n+1} \right)^2 \\ &= \frac{\theta^2}{(2n+1)^2} \left( \frac{n}{n+1} + 1 \right) \\ &= \frac{\theta^2}{(n+1)(2n+1)} \end{aligned} \tag{1}$$

Since

$$\frac{n}{n+1} \rightarrow 1$$

as  $n \rightarrow \infty$ , the squared bias and the variance contribute equally to the MSE. □

- (f) Find a method of moments estimator for  $\theta$  based on the  $X_i$  and call this  $\hat{\theta}_X$ .

*Solution.* The marginal density of  $X_i$  is given by

$$f_X(x) = \frac{2\sqrt{\theta^2 - x^2}}{\theta^2\pi}$$

which is symmetric about the origin, so  $\mu_1 = E[X_i] = 0$ . Then

$$\mu_2 = E[X_i^2] = \int_{-\theta}^{\theta} x^2 \cdot \frac{2\sqrt{\theta^2 - x^2}}{\theta^2\pi} dx = \frac{\theta^2}{4}$$

according to Wolfram, so the method of moments estimate is

$$\hat{\theta}_x = 2\sqrt{\hat{\mu}_2}.$$

□

- (g) Compare the performance of the MLE and the method of moments estimator as follows: In R, do the following 10000 times. Sample the uniform distribution in the unit circle using a sample of size 10, and compute the three estimators (MLE, MoM  $X_i$ , MoM  $Y_i$ ). Compute estimates of the bias, the variance, and the MSE of each. Estimate the correlation coefficient between  $\hat{\theta}_x$  and  $\hat{\theta}_y$ . Assuming your estimate in the previous parts are correct, how much should we improve the variance of one of  $\hat{\theta}_x$  or  $\hat{\theta}_y$  by averaging them?
- (h) Show that for the method of moments estimator and the MLE, is it the case that the distribution of  $\hat{\theta}/\theta$  does not depend on  $\theta$ . Explain why this means we can write

$$MSE_{\theta}(\hat{\theta}) = \theta^2 \left( MSE_{\theta=1}(\hat{\theta}) \right)$$

From this, explain why it suffices that we compare the two estimators when  $\theta = 1$ .

## Chapter 9: Testing Hypotheses and Assessing Goodness of Fit

2. Which of the following hypotheses are simple, and which are composite?
- $X$  follows a uniform distribution on  $[0, 1]$ .
  - A die is unbiased.
  - $X$  follows a normal distribution with mean 0 and variance  $\sigma^2 > 10$ .
  - $X$  follows a normal distribution with mean  $\mu = 0$ .
5. True or false, and state why:
- The significance level of a statistical test is equal to the probability that the null hypothesis is true.
  - If the significance level of a test is decreased, the power would be expected to increase.
  - If a test is rejected at the significance level  $\alpha$ , the probability that the null hypothesis is true equals  $\alpha$ .
  - The probability that the null hypothesis is falsely rejected is equal to the power of the test.
  - A type I error occurs when the test statistic falls in the rejection region of the test.
  - A type II error is more serious than a type I error.
  - The power of a test is determined by the null distribution of the test statistic.
  - The likelihood ratio is a random variable.
4. Let  $X$  have one of the following distributions:

$X$	$H_0$	$H_A$
$x_1$	0.2	0.1
$x_2$	0.3	0.4
$x_3$	0.3	0.1
$x_4$	0.2	0.4

- Compare the likelihood ratio,  $\Lambda$ , for each possible value  $X$  and order the  $x_i$  according to  $\Lambda$ .
- What is the likelihood ratio test of  $H_0$  versus  $H_A$  at the level  $\alpha = 0.2$ ? What is the test at the level  $\alpha = 0.5$ ?
- If the prior probabilities are  $P(H_0) = P(H_A)$ , which outcomes favor  $H_0$ ?
- What prior probabilities correspond to the decision rules with  $\alpha = 0.2$  and  $\alpha = 0.5$ ?

7. Let  $X_1, \dots, X_n$  be a sample from a Poisson distribution. Find the likelihood ratio for testing  $H_0 : \lambda = \lambda_0$  versus  $H_a : \lambda = \lambda_1$ , where  $\lambda_1 > \lambda_0$ . Use the fact that the sum of independent Poisson random variables follows a Poisson distribution to explain how to determine a rejection region for a test at level  $\alpha$ .
9. Let  $X_1, \dots, X_{25}$  be a sample from a normal distribution having a variance of 100. Find the rejection region for a test at level  $\alpha = 0.10$  of  $H_0 : \mu = 0$  versus  $H_A : \mu = 1.5$ . What is the power of the test? Repeat for  $\alpha = 0.01$ .