Statistics X_400004 Resit 05.02.2019 Solutions

Exercise 1 [19 points (Exam 1 resit); 14 points (Full resit)]
Let $X_1, \ldots, X_n$ be an i.i.d. sample from the Uniform$[\theta, 2\theta]$ distribution (see Appendix 1), where $\theta > 0$ is an unknown parameter.

Hint: To answer the questions below, you can use without derivation (the relevant ones among) the following facts: $X_i = \theta + \theta Y_i = \theta(1 + Y_i)$, where the Y_i's $\overset{\text{i.i.d.}}{\sim}$ Uniform$[0, 1]$. In particular, $X_{(1)} = \theta(1 + Y_{(1)})$ and $X_{(n)} = \theta(1 + Y_{(n)})$. Furthermore, $\mathbb{E}Y_{(1)} = \dfrac{1}{n+1}$, $\mathbb{E}Y_{(n)} = \dfrac{n}{n+1}$, and $\mathbb{V}Y_{(1)} = \mathbb{V}Y_{(n)} = \dfrac{n}{(n+1)^2(n+1)}$.

(a) [6 points] Give the moment estimator $\widehat{\theta}_{\text{MM}}$ and the maximum likelihood estimator $\widehat{\theta}_{\text{ML}}$ of the parameter θ.

By Appendix 1, $\mathbb{E}X_1 = \dfrac{\theta + 2\theta}{2} = \dfrac{3}{2}\theta$. We have $\theta = \dfrac{2}{3}\mathbb{E}X_1$, and the moment estimator is given by

$$\widehat{\theta}_{\text{MM}} = \frac{2}{3}\overline{X}.$$

Now we find the maximum-likelihood estimator for θ. The p.d.f. of the Uniform$[\theta, 2\theta]$ distribution is $p_\theta(x) = \dfrac{1}{2\theta - \theta} = \dfrac{1}{\theta}$, and hence the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} p_\theta(X_i) = \frac{1}{\theta^n},$$

which is decreasing in θ. There are restrictions on θ from the data: since $\theta \leq X_i \leq 2\theta$ for all i, we have $\theta \leq X_{(1)}$ and $X_{(n)} \leq 2\theta$. I.e. the feasible values of θ are

$$\frac{X_{(n)}}{2} \leq \theta \leq X_{(1)}.$$

Since the likelihood $L(\theta)$ is decreasing, it is maximized by the smallest feasible θ, i.e.

$$\widehat{\theta}_{\text{ML}} = \text{argmax}_{\frac{X_{(n)}}{2} \leq \theta \leq X_{(1)}} L(\theta) = \frac{X_{(n)}}{2}.$$

(b) [5 points (Exam 1 resit only)] Transform the estimators $\widehat{\theta}_{\text{MM}}$ and $\widehat{\theta}_{\text{ML}}$ that you found in part *(a)* into unbiased estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ of the parameter θ. (If $\widehat{\theta}_{\text{MM}}$ is already unbiased, just put $\widehat{\theta}_1 = \widehat{\theta}_{\text{MM}}$. If $\widehat{\theta}_{\text{ML}}$ is already unbiased, just put $\widehat{\theta}_2 = \widehat{\theta}_{\text{ML}}$.)

We have

$$\mathbb{E}\widehat{\theta}_{\text{MM}} = \frac{2}{3}\mathbb{E}\overline{X} = \frac{2}{3}\mathbb{E}X_1 \overset{(a)}{=} \frac{2}{3} \cdot \frac{3}{2}\theta = \theta.$$

That is, $\widehat{\theta}_{\text{MM}}$ is an unbiased estimator for θ and we put $\widehat{\theta}_1 = \dfrac{2}{3}\overline{X}$.
By the hint,

$$\mathbb{E}\widehat{\theta}_{\text{ML}} = \frac{\mathbb{E}X_{(n)}}{2} = \frac{\theta(1 + \mathbb{E}Y_{(n)})}{2} = \frac{\theta}{2}\Big(1 + \frac{n}{n+1}\Big) = \frac{\theta}{2} \cdot \frac{n+1+n}{n+1} = \frac{2n+1}{2(n+1)}\theta.$$

That is, $\widehat{\theta}_{\text{ML}}$ is biased. We can make it unbiased by rescaling: put

$$\widehat{\theta}_2 = \frac{2(n+1)}{2n+1}\widehat{\theta}_{\text{ML}} = \frac{2(n+1)}{2n+1}\frac{X_{(n)}}{2} = \frac{n+1}{2n+1}X_{(n)}.$$

Then

$$\mathbb{E}\widehat{\theta}_2 = \frac{2(n+1)}{2n+1}\mathbb{E}\widehat{\theta}_{\mathrm{ML}} = \frac{2(n+1)}{2n+1}\cdot\frac{2n+1}{2(n+1)}\theta = \theta,$$

i.e. $\widehat{\theta}_2 = \frac{n+1}{2n+1}X_{(n)}$ is an unbiased estimator for θ.

(c) [8 points] Which of the unbiased estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ that you found in part *(b)* is better for large data sets? (That is, as $n \to \infty$. You do not have to specify for which n exactly one estimator is better than the other).

Hint: If you do not have an answer to part *(b)*, compare $\hat{\theta}_1 = \frac{2}{3}\overline{X}$ and $\hat{\theta}_2 = \frac{n+1}{2n+1}X_{(n)}$ under the assumption that these are unbiased estimators for θ.

Since $\widehat{\theta}_1$ is unbiased, we have

$$\mathrm{MSE}(\widehat{\theta}_1) = \mathbb{V}\widehat{\theta}_1 = \mathbb{V}\left(\frac{2}{3}\overline{X}\right) = \frac{4}{9}\mathbb{V}\overline{X} = \frac{4}{9n}\mathbb{V}X_1 \overset{\text{Appendix 1}}{=} \frac{4}{9n}\frac{(2\theta-\theta)^2}{12} = \frac{4}{9n}\frac{\theta^2}{12} = \frac{1}{27n}\theta^2.$$

Since $\widehat{\theta}_2$ is unbiased as well,

$$\mathrm{MSE}(\widehat{\theta}_2) = \mathbb{V}\widehat{\theta}_2 = \mathbb{V}\left(\frac{n+1}{2n+1}X_{(n)}\right) = \left(\frac{n+1}{2n+1}\right)^2\mathbb{V}X_{(n)},$$

where

$$\mathbb{V}X_{(n)} = \mathbb{V}\Big(\theta(1+Y_n)\Big) = \theta^2\mathbb{V}(1+Y_{(n)}) = \theta^2\mathbb{V}(Y_n) \overset{\text{Hint}}{=} \frac{n}{(n+1)^2(n+2)}\theta^2.$$

Hence,

$$\mathrm{MSE}(\widehat{\theta}_2) = \frac{(n+1)^2}{(2n+1)^2}\mathbb{V}X_{(n)} = \frac{(n+1)^2}{(2n+1)^2}\cdot\frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{n}{(2n+1)^2(n+2)}\theta^2.$$

As $n \to \infty$, $\mathrm{MSE}(\widehat{\theta}_2) \approx \frac{1}{4n^2}\theta^2$ decays faster (becomes smaller) than $\mathrm{MSE}(\widehat{\theta}_1) = \frac{1}{27n}\theta^2$. Hence, for large data sets, the estimator $\widehat{\theta}_2$ is better than $\widehat{\theta}_1$.

Exercise 2 [7 points]

Let $X_1, \ldots, X_n$ be an i.i.d. sample from the Geometric(θ) distribution (see Appendix 1) with unknown parameter $\theta \in (0,1)$. Give the Bayes estimator for θ if the prior belief about θ is the Beta(3,2) distribution (see Appendix 1).

Hint: The posterior is proportional to the product of the prior and the likelihood. The posterior distribution is a common distribution (see Appendix 1).

By Appendix 1, the prior density is

$$\pi(\theta) \propto \theta^{3-1}(1-\theta)^{2-1}.$$

The likelihood is

$$L(\theta) = \prod_{i=1}^{n} p_\theta(X_i) = \prod_{i=1}^{n}\Big((1-\theta)^{X_i-1}\theta\Big) = (1-\theta)^{\sum_{i=1}^{n}(X_i-1)}\theta^n = (1-\theta)^{n\overline{X}-n}\theta^n.$$

The posterior on θ is proportional to the prior $\times$ the likelihood,

$$p_{\overline{\theta}|X_1,\ldots,X_n}(\theta) \propto \pi(\theta) \times L(\theta)$$

$$\propto [\theta^{3-1}(1-\theta)^{2-1}] \times [(1-\theta)^{n\overline{X}-n}\theta^n] = \theta^{\overbrace{(n+3)}^{=:\,\alpha_{\text{new}}}-1}(1-\theta)^{\overbrace{(n\overline{X}-n+2)}^{=:\,\beta_{\text{new}}}-1}.$$

By Appendix 1, the posterior belief about θ is $\overline{\theta}|X_1,\ldots,X_n \sim \text{Beta}(\alpha_{\text{new}},\beta_{\text{new}})$, and the Bayes estimator for θ is

$$\widehat{\theta}_{\text{B}} = \mathbb{E}[\overline{\theta}|X_1,\ldots,X_n] = \mathbb{E}\,\text{Beta}(\alpha_{\text{new}},\beta_{\text{new}}) = \frac{\alpha_{\text{new}}}{\alpha_{\text{new}}+\beta_{\text{new}}} = \frac{n+3}{n\overline{X}+2}.$$

Exercise 3 [19 points (Exam 1 resit only)]

Until the last schedule adjustment, the expected number of passengers that used a particular NS-station per working day was 1200. After the schedule adjustment, the station feels quieter. The passenger numbers $X_1,\ldots,X_4$ for 4 consecutive working days since the schedule adjustment average to $\overline{X} = 1090$. Do these data indicate that the number of people using the station on working days has indeed dropped? Assume normal distribution and a known standard deviation of 100.

(a) [3 points] Formulate an appropriate statistical model and a null and alternative hypotheses.

The statistical model is

$$X_1,\ldots,X_n \overset{\text{i.i.d.}}{\sim} N(\mu,\sigma^2), \quad \text{where } n=4 \text{ and } \sigma=100 \text{ is known.}$$

We will test the hypotheses

$$H_0\colon \mu \geq 1200 \quad \text{VS} \quad H_1\colon \mu < 1200.$$

(We are looking for evidence of decrease and therefore put "after $<$ before" in H_1.)

(b) [10 points] Test the hypotheses from part *(a)*. Allow for a 2.5% chance of type-1 error. Report the test statistic, its distribution on the border of H_0, the critical region, the $p-$value, and the conclusion.

We will use the Gauss test. I.e. the the test statistic and its distribution on the border of H_0 are

$$T := \frac{\overline{X}-1200}{\sigma/\sqrt{n}} = \frac{\overline{X}-1200}{100/\sqrt{4}} \overset{\mu=1200}{\sim} N(0,1).$$

Allowing for a 2.5% chance of type-1 error means testing at significance level $\alpha = 0.025$. Then the critical region is (H_0 is rejected if)

$$T \leq -z_\alpha, \quad \text{where } z_\alpha = z_{0.025} \overset{\text{Appendix 3}}{=} 1.96.$$

We observe $\overline{X} = 1090$. As or more unusual under $H_0\colon \mu \geq 1200$ would be to observe $\overline{X} \leq 1090$ (even smaller values). Hence,

$$\begin{aligned} p\text{-value} &= \max_{\mu\geq 1200} \mathbb{P}(\overline{X} \leq 1090) = \max_{\mu\geq 1200} \mathbb{P}(N(\mu,100^2/4) \leq 1090) = \mathbb{P}(N(1200,100^2/4) \leq 1090) \\ &= \mathbb{P}(N(0,1) \leq \frac{1090-1200}{100/\sqrt{4}}) = \mathbb{P}(N(0,1) \leq -2.2) = 1-\Phi(2.2) \overset{\text{Appendix 3}}{=} 0.0139. \end{aligned}$$

Since

$$(p\text{-value} = 0.0139) < (\alpha = 0.025),$$

we reject H_0. That is, the number of passengers using the station has indeed dropped after the schedule adjustment.

(c) [6 points] Assume the true expected passenger number since the schedule adjustment is 1100. For how many days should the passenger numbers be observed so that the power of the test from *(b)* is at least 90%?

As mentioned in *(b)*, if there are n observations, then the critical region is (H_0 is rejected if)

$$\frac{\overline{X} - 1200}{100/\sqrt{n}} \le -1.96.$$

The power of the test when $\mu = 1100$ is given by

$$\beta(1100) := \mathbb{P}_{\mu=1100}(\text{reject (wrong) } H_0) = \mathbb{P}_{\mu=1100}\Big(\frac{\overline{X} - 1200}{100/\sqrt{n}} \le -1.96\Big)$$
$$= \mathbb{P}_{\mu=1100}\Big(\underbrace{\frac{\overline{X} - 1100}{100/\sqrt{n}}}_{\sim N(0,1)} - \frac{100}{100/\sqrt{n}} \le -1.96\Big) = \mathbb{P}(N(0,1) \le \sqrt{n} - 1.96).$$

Then $\beta(1100) \ge 0.9$ is equivalent to

$$\sqrt{n} - 1.96 \ge z_{0.1} \overset{\text{Appendix 3}}{=} 1.28$$
$$\Leftrightarrow n \ge (1.96 + 1.28)^2$$
$$\Leftrightarrow n \ge 11.$$

Exercise 4 [12 points]

A cab company is choosing between two tire brands, A and B. Previously, the company bought 9 tires by each brand and put them to a wear test. More specifically, it put new back tires on 9 cabs, one back tire by brand A and the other back tire by brand B, and recorded the numbers $X_1, \dots, X_9$ of thousands driven kilometres the test tires by brand A had lasted, and the numbers $Y_1, \dots, Y_9$ of thousands driven kilometres the test tires by brand B had lasted. Some of the data summaries are

$$\overline{X} = 59, \quad S_X^2 = 30.8, \quad \overline{Y} = 57.5, \quad S_Y^2 = 33.6,$$
$$S_Z^2 = 2.8, \quad \text{where } Z_i = X_i - Y_i.$$

Do the data present evidence that, as for wear, one tire brand is better than the other? Carry out a suitable statistical test at significance level 0.05. Report

(a) [3 points] the statistical model and the null and alternative hypotheses,

The experiment design is paired, the statistical model is

$$Z_1, \dots, Z_n \overset{\text{i.i.d.}}{\sim} N(\Delta, \sigma^2), \quad \text{where } n = 9 \text{ and both } \Delta \text{ and } \sigma \text{ are unknown.}$$

We will test the hypotheses

$$H_0\colon \Delta = 0 \quad \text{VS} \quad H_1\colon \Delta \neq 0.$$

(We are looking for evidence of any difference and therefore put "$\Delta \neq 0$" in H_1.)

(b) [3 points] the test statistic and its distribution when, on average, there is no difference between the two tire brands as for wear,

We will use the paired t-test. I.e. the test statistic and its distribution in case, on average, there is no difference are

$$T := \frac{\overline{Z} - 0}{S_Z/\sqrt{n}} \overset{\Delta=0}{\sim} t_{n-1}.$$

(c) [6 points] the critical region and the conclusion.

We reject $H_0\colon \Delta = 0$ at significance level $\alpha = 0.05$ if

$$|T| \geq (t_{n-1,1-\alpha/2} = t_{8,0.975} \overset{\text{Appendix 4}}{=} 2.31).$$

We observe

$$T = \frac{\overline{Z}}{S_Z/\sqrt{n}} = \frac{\overline{X} - \overline{Y}}{S_Z/\sqrt{n}} = \frac{59 - 57.5}{\sqrt{2.8}/\sqrt{9}} = 2.69.$$

Since $|2.69| > 2.31$, we reject H_0. That is, the data do present evidence that there is difference between the two tire brands as for wear (brand A is better).

Exercise 5 [13 points (Exam 2 resit); 10 points (Full resit)]

Let $X_1, \dots, X_{10}$ be an i.i.d. sample from the normal $N(0, \sigma^2)$ distribution, and $Y_1, \dots, Y_5$ an i.i.d. sample from the normal $N(0, 4\sigma^2)$ distribution. The two samples are independent. The parameter $\sigma^2 > 0$ is unknown.

(a) [3 points] Construct a pivot out of σ^2 and the sample $X_1, \dots, X_{10}$. What is the distribution of this pivot?

By standartization,

$$\frac{X_1 - 0}{\sigma}, \dots, \frac{X_{10} - 0}{\sigma} \overset{\text{i.i.d.}}{\sim} N(0,1),$$

and, by definition of the χ^2 distribution,

$$\sum_{i=1}^{10} \left(\frac{X_i - 0}{\sigma} \right)^2 = \underbrace{\frac{10\overline{X^2}}{\sigma^2}}_{\text{pivot}} \sim \chi^2_{10}.$$

(b) [3 points (Exam 2 resit only)] Construct a pivot out of σ^2 and both samples. What is the distribution of this pivot?

By standartization,

$$\frac{X_1 - 0}{\sigma}, \dots, \frac{X_{10} - 0}{\sigma}, \frac{Y_1 - 0}{2\sigma}, \dots, \frac{Y_5 - 0}{2\sigma} \overset{\text{i.i.d.}}{\sim} N(0,1),$$

and, by definition of the χ^2 distribution,

$$\sum_{i=1}^{10} \left(\frac{X_i - 0}{\sigma} \right)^2 + \sum_{j=1}^{5} \left(\frac{Y_j - 0}{2\sigma} \right)^2 = \frac{10\overline{X^2}}{\sigma^2} + \frac{5\overline{Y^2}}{4\sigma^2} = \underbrace{\frac{1}{\sigma^2}\left(10\overline{X^2} + \frac{5\overline{Y^2}}{4} \right)}_{\text{pivot}} \sim \chi^2_{15}.$$

(c) [3 points] Based on either of the pivots from *(a)* or *(b)* (only one of them), construct a confidence interval for σ^2 of confidence level 0.99.

Using the pivot from *(b)*, we have

$$\mathbb{P}\left\{ \chi^2_{15,0.005} \leq \underbrace{\frac{1}{\sigma^2}\left(10\overline{X^2} + \frac{5\overline{Y^2}}{4} \right)}_{\text{pivot} \sim \chi^2_{15}} \leq \chi^2_{15,0.995} \right\} = 0.99.$$

which is equivalent to

$$\underbrace{\mathbb{P}\left\{ \frac{10\overline{X^2} + 5\overline{Y^2}/4}{\chi^2_{15,0.995}} \leq \sigma^2 \leq \frac{10\overline{X^2} + 5\overline{Y^2}/4}{\chi^2_{15,0.005}} \right\}}_{\text{CI for } \sigma^2 \text{ of level } 0.99} = 0.99.$$

(d) [4 points] It has been observed that $\overline{X^2} = 0.7$ and $\overline{Y^2} = 1.2$. Test whether σ^2 deviates from 1 using the confidence interval from *(c)*. What is the significance level of this test?

We can test $H_0 : \sigma^2 = 1$ VS $H_1 : \sigma^2 \neq 1$ at significance level $\alpha = 0.01$ as follows:
if $1 \notin$ CI for σ^2 of confidence level $1 - \alpha = 0.99$, then reject H_0;
if $1 \in$ the CI, then fail to reject H_0.

With $\overline{X} = 0.4$ and $\overline{Y^2} = 1.2$ observed, the CI of level 0.99 from *(c)* is

$$\left[\frac{10 \cdot 0.4 + 5 \cdot 1.2/4}{32.801} \leq \sigma^2 \leq \frac{10 \cdot 0.4 + 5 \cdot 1.2/4}{4.601}\right] = [0.168, 1.195].$$

Since $1 \in [0.168, 1.195]$, we fail reject $H_0 \colon \sigma^2 = 1$ at significance level 0.01. I.e. no evidence that σ^2 deviates from 1.

Exercise 6 [20 points]

In parts *(a)* **and** *(b)*, $X_1, \ldots, X_n$ is an i.i.d. sample from the Poisson(λ) distribution with unknown parameter $\lambda > 0$ (see Appendix 1).

(a) [7 points] The maximum likelihood estimator for λ is $\widehat{\lambda}_{\text{ML}} = \overline{X}$, and i_λ denotes the Fisher information (see Appendix 2). For large n, $\sqrt{n i_\lambda}(\widehat{\lambda}_{\text{ML}} - \lambda) \approx N(0, 1)$. Construct a confidence interval for λ of an approximate confidence level $1 - \alpha$.

In the near-pivot given, we estimate the Fisher information (see below), i.e. we are using the near-pivot

$$\sqrt{n\widehat{i_\lambda}}(\widehat{\lambda}_{\text{ML}} - \lambda) \approx N(0, 1).$$

From the above pivot, the Wald CI for λ of confidence level $\approx 1 - \alpha$ follows,

$$\lambda = \widehat{\lambda}_{\text{ML}} \pm \frac{1}{\sqrt{n\widehat{i_\lambda}}} z_{\alpha/2}.$$

Now we compute/estimate the Fisher information: using Appendix 1 and 2,

$$\begin{aligned}
\ell_\lambda(k) &= \log p_\lambda(k) = \log(e^{-\lambda}\frac{\lambda^k}{k!}) = -\lambda + k\log\lambda - \log(k!),\\
\dot{\ell}_\lambda(x) &= \frac{\partial \ell}{\partial \lambda}\ell_\lambda(x) = -1 + \frac{k}{\lambda},\\
\ddot{\ell}_\lambda(x) &= \frac{\partial^2 \ell}{\partial \lambda^2}\ell_\lambda(x) = -\frac{k}{\lambda^2},
\end{aligned}$$

and hence,

$$i_\lambda = -\mathbb{E}\ddot{\ell}_\lambda(X_1) = \frac{\mathbb{E}X_1}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$

We use the plug-in estimator for the Fisher information,

$$\widehat{i_\lambda} = \frac{1}{\widehat{\lambda}_{\text{ML}}} = \frac{1}{\overline{X}}.$$

So the Wald CI for λ of approximate confidence level $1 - \alpha$ is

$$\lambda = \overline{X} \pm \sqrt{\frac{\overline{X}}{n}} z_{\alpha/2}.$$

(b) [5 points] Does the Cramèr-Rao lower bound (see Appendix 2) imply that $\overline{X}$ is a UMVU estimator for λ?

First of all, note that $\overline{X}$ (sample mean) is an unbiased estimator for λ (population mean), indeed

$$\mathbb{E}\overline{X} \overset{\text{i.i.d.}}{=} \mathbb{E}X_1 = \lambda.$$

Now we compute the variance of $\overline{X}$ and compare it to the Cramèr-Rao lower bound (CRLB). We have

$$\mathbb{V}\overline{X} = \frac{\mathbb{V}X_1}{n} = \frac{\lambda}{n}.$$

Since we are estimating λ, we compute the CRLB with $g(\lambda) = \lambda$ and $i_\lambda \overset{(a)}{=} \dfrac{1}{\lambda}$,

$$\text{CRLB} = \frac{(g'(\lambda))^2}{ni_\lambda} = \frac{1}{n/\lambda} = \frac{\lambda}{n}.$$

We have

$$\mathbb{V}\overline{X} = \text{CRLB},$$

while $\mathbb{V}\widehat{\lambda} \geq \text{CRLB}$ for any other unbiased estimator $\widehat{\lambda}$ of λ. Hence the Cramèr-Rao lower bound does imply that $\overline{X}$ is an UMVU estimator for λ.

In part *(c)*, $X_1, \ldots, X_n$ is an i.i.d. sample the normal $N(0, \sigma^2)$ distribution with unknown variance $\sigma^2 > 0$ (see Appendix 1).

(c) [8 points] Is $\overline{X^2}$ a sufficient and complete statistic (see Appendix 2)? Is $\overline{X^2}$ a UMVU estimator for σ^2?

We have

$$L(\sigma^2) \overset{\text{i.i.d.}}{=} \prod_{i=1}^n p_{0,\sigma^2}(X_i) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{X_i^2}{2\sigma^2}} \right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\sum_{i=1}^n \frac{X_i^2}{2\sigma^2}}$$

$$= \overbrace{\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n}^{c(\sigma^2)} e^{\overbrace{-\frac{n}{2\sigma^2}}^{Q_1(\sigma^2)} \cdot \overbrace{\overline{X^2}}^{V_1}} \cdot \overbrace{1}^{h}.$$

By Appendix 2, the statistical model forms a 1-dimensional exponential family. The statistic $V_1 = \overline{X^2}$ is sufficient.

Since the set $\{Q_1(\sigma^2) \colon \sigma^2 > 0\} = \left\{ -\dfrac{n}{2\sigma^2} \colon \sigma^2 > 0 \right\} = (-\infty, 0)$ has interior points in $\mathbb{R}^1$ (contains an interval), the statistic $V_1 = \overline{X^2}$ is also complete.

Now consider the estimator $\overline{X^2}$ for σ^2. This is an unbiased estimator for σ^2 because

$$\mathbb{E}\overline{X^2} \overset{\text{i.i.d.}}{=} \mathbb{E}X_1^2 = \mathbb{V}X_1 + (\mathbb{E}X_1)^2 = \sigma^2 + 0^2 = \sigma^2.$$

Finally, the estimator $\overline{X^2}$ is unbiased for σ^2 and is a sufficient and complete statistic (made out of the sufficient and complete statistic $\overline{X^2}$). Hence, $\overline{X^2}$ is a UMVU estimator for σ^2.

Appendix 1: Some relevant distributions

Uniform$[a, b]$ distribution, $a < b$

The p.d.f. is given by $\begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$

The expectation is $\frac{a+b}{2}$, the variance is $\frac{(b-a)^2}{12}$.

Geometric(θ) distribution, $\theta \in (0, 1]$

The p.m.f. is given by $(1-\theta)^{k-1}\theta$, $k = 1, 2, \ldots$

The expectation is $\frac{1}{\theta}$, the variance is $\frac{1-\theta}{\theta^2}$.

Beta$(\alpha$,$\beta)$ distribution, $\alpha > 0$, $\beta > 0$

The p.d.f. is proportional to $\begin{cases} \theta^{\alpha-1}(1-\theta)^{\beta-1}, & 0 < \theta < 1, \\ 0, & \text{otherwise.} \end{cases}$

The expectation is $\frac{\alpha}{\alpha+\beta}$, the variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Poisson(λ) distribution, $\lambda > 0$

The p.m.f. is given by $e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$

The expectation is λ, the variance is λ.

Normal$(\mu$,$\sigma^2)$ distribution, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

The p.d.f. is given by $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$

The expectation is μ, the variance is σ^2.

Appendix 2: Some relevant facts from optimality theory

Fisher information Let $X_1, \ldots, X_n$ be i.i.d. with marginal p.d.f./p.m.f. $p_\theta(x)$ and $\ell_\theta(x) := \log p_\theta(x)$, $\dot{\ell}_\theta(x) := \frac{\partial}{\partial\theta}\ell_\theta(x)$, $\ddot{\ell}_\theta(x) := \frac{\partial^2}{\partial\theta^2}\ell_\theta(x)$. Then the Fisher information is given by

$$i_\theta := \mathbb{V}\ \dot{\ell}_\theta(X_1) = -\mathbb{E}\ \ddot{\ell}_\theta(X_1).$$

Cramer-Rao lower bound Let $X_1, \ldots, X_n$ be i.i.d. random variables from a distribution parametrized by θ. Under certain conditions, every unbiased estimator $\widehat{g(\theta)}$ for $g(\theta)$ satisfies

$$\mathbb{V}\ \widehat{g(\theta)} \geq \frac{(g'(\theta))^2}{n i_\theta}.$$

Exponential family A statistical model parametrized by θ forms a k-dimensional exponential family if the likelihood is of the form

$$L(\theta) = c(\theta) \cdot e^{\sum_{j=1}^{k} Q_j(\theta) \cdot V_j(X_1, \ldots, X_n)} \cdot h(X_1, \ldots, X_n).$$

.

Appendix 3: Table normal distribution

	0	1	2	3	4	5	6	7	8	9
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999

Table 1: Distribution function of the standard normal distribution on the interval $[1.1, 3.09]$. The value in the table is $\Phi(x)$ for $x = a + b/100$ where a indicates the row and b indicates the column.

Appendix 4: Table t-distribution

df	0.6	0.7	0.75	0.8	0.85	0.9	0.925	0.95	0.975	0.98	0.99	0.999
8	0.26	0.55	0.71	0.89	1.11	1.4	1.59	1.86	2.31	2.45	2.9	4.5
9	0.26	0.54	0.7	0.88	1.1	1.38	1.57	1.83	2.26	2.4	2.82	4.3
10	0.26	0.54	0.7	0.88	1.09	1.37	1.56	1.81	2.23	2.36	2.76	4.14
11	0.26	0.54	0.7	0.88	1.09	1.36	1.55	1.8	2.2	2.33	2.72	4.02
12	0.26	0.54	0.7	0.87	1.08	1.36	1.54	1.78	2.18	2.3	2.68	3.93
13	0.26	0.54	0.69	0.87	1.08	1.35	1.53	1.77	2.16	2.28	2.65	3.85
14	0.26	0.54	0.69	0.87	1.08	1.35	1.52	1.76	2.14	2.26	2.62	3.79
15	0.26	0.54	0.69	0.87	1.07	1.34	1.52	1.75	2.13	2.25	2.6	3.73
16	0.26	0.54	0.69	0.86	1.07	1.34	1.51	1.75	2.12	2.24	2.58	3.69
17	0.26	0.53	0.69	0.86	1.07	1.33	1.51	1.74	2.11	2.22	2.57	3.65
18	0.26	0.53	0.69	0.86	1.07	1.33	1.5	1.73	2.1	2.21	2.55	3.61
19	0.26	0.53	0.69	0.86	1.07	1.33	1.5	1.73	2.09	2.2	2.54	3.58
20	0.26	0.53	0.69	0.86	1.06	1.33	1.5	1.72	2.09	2.2	2.53	3.55
21	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.72	2.08	2.19	2.52	3.53
22	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.72	2.07	2.18	2.51	3.5
23	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.71	2.07	2.18	2.5	3.48
24	0.26	0.53	0.68	0.86	1.06	1.32	1.49	1.71	2.06	2.17	2.49	3.47
25	0.26	0.53	0.68	0.86	1.06	1.32	1.49	1.71	2.06	2.17	2.49	3.45
26	0.26	0.53	0.68	0.86	1.06	1.31	1.48	1.71	2.06	2.16	2.48	3.43
27	0.26	0.53	0.68	0.86	1.06	1.31	1.48	1.7	2.05	2.16	2.47	3.42

Table 2: Quantiles (columns) of the t-distribution with 8 to 27 degrees of freedom (rows).

Appendix 5: Table Chi-square distribution

df	0.001	0.005	0.01	0.025	0.05	0.1	0.125	0.2	0.25	0.333	0.5
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337
21	6.447	8.034	8.897	10.283	11.591	13.240	13.873	15.445	16.344	17.720	20.337
22	6.983	8.643	9.542	10.982	12.338	14.041	14.695	16.314	17.240	18.653	21.337
23	7.529	9.260	10.196	11.689	13.091	14.848	15.521	17.187	18.137	19.587	22.337
24	8.085	9.886	10.856	12.401	13.848	15.659	16.351	18.062	19.037	20.523	23.337
25	8.649	10.520	11.524	13.120	14.611	16.473	17.184	18.940	19.939	21.461	24.337

df	0.6	0.667	0.75	0.8	0.87	0.9	0.95	0.975	0.99	0.995	0.999
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620

Table 3: Quantiles (columns) of the chi-square distribution with 6 to 25 degrees of freedom (rows).