Math 357 Final Review List

By Johnson Zhang

This document may contain potential typos and mistakes, please let the author know once you find it. Contact information: ***zhangjohnson729gmail.com***

Contents

1 Basic Results

1.1 Basic Properties of Random Samples

Below, denote $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ as a random sample. Then we know that

1. $\overline{X}_n \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$;

2. $\overline{X}_n \perp\!\!\!\perp S_n^2$, and indeed $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$;

The t distribution can be obtained by a standard normal distribution $Z \sim N(0,1)$ and a $V \sim \chi^2_\nu$ distribution, where

$$T = \frac{Z}{\sqrt{V/\nu}} \overset{\text{pdf}}{=} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right).$$

Recall from CLT, where in a random sample $X_1, \cdots, X_n$ with $\mathbb{E}X = \mu, Var(X) = \sigma^2$, we have

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

We are also able to obtain T by a variation of central limit theorem, in a random sample $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$, we have

$$T = \frac{\overline{X}_n - \mu}{\sqrt{S_n^2/n}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t(n-1)$$

The F distribution can be viewed as the ratio of two χ-squared distribution, where

$$F(m,n) \sim \frac{\chi^2_m/m}{\chi^2_n/n}$$

where if we consider two mutually independent random samples $X_1, \cdots, X_m \sim N(\mu_1, \sigma_1^2)$ and $Y_1, \cdots, Y_n \sim N(\mu_2, \sigma_2^2)$, then

$$F = \frac{S_m^2/\sigma_1^2}{S_n^2/\sigma_n^2} \sim F(m-1, n-1).$$

1.2 Transformations Between Distributions

Below, the independence of random variables is assumed.

1. If $X \sim Exp(\theta)$ where $f(x) = \theta e^{-\theta x}$, then $2\theta X \sim \chi^2_2$, where the square of a standard normal is χ^2_1, and $\chi^2_a + \chi^2_b = \chi^2_{a+b}$.

2. If $X \sim Exp(\theta)$ where $f(x) = \theta e^{-\theta x}$, then $X_1 + \cdots + X_n \sim Gamma(n, \theta)$.

3. (Cochran's Theorem) Suppose we have $Z_1, \cdots, Z_n \overset{i.i.d}{\sim} N(0,1)$, then

$$\sum_{i=1}^{n} (Z_i - \overline{Z}_n)^2 \sim \chi^2_{n-1}$$

where $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \sim N\left(0, \frac{1}{n}\right)$.

4. The sample mean of a χ^2 distribution family is distributed according to a Gamma distribution. That is, if $X_1, \cdots, X_n \overset{i.i.d}{\sim} \chi_k^2$, then

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim Gamma\left(\frac{nk}{2}, \frac{2}{n}\right)$$

where $Gamma(\alpha, \beta) \sim \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$. That is, we have the following relations between Gamma distribution and Chi-squared distribution: If $Q \sim \chi_\nu^2$ and $c > 0$, then $cQ \sim Gamma\left(\frac{\nu}{2}, 2c\right)$.

5. If $X \sim \chi_a^2$ and $Y \sim \chi_b^2$, then $\frac{X}{X+Y} \sim Beta\left(\frac{a}{2}, \frac{b}{2}\right)$, a similar result is also used for the ratio of Gamma distributions: If $X \sim Gamma(\alpha, \theta)$ and $Y \sim Gamma(\beta, \theta)$, then $\frac{X}{X+Y} \sim Beta(\alpha, \beta)$. Furthermore, using **Basu's Theorem**, $\frac{X}{X+Y} \perp\!\!\!\perp X+Y$, and $X+Y \sim Gamma(\alpha+\beta, \theta)$.

6. Let $X \sim Uniform(0, 1)$, then $-2\log X \sim \chi_2^2$.

7. (Poisson Process with Gamma Distribution) If $X \sim Gamma(\alpha, \beta)$ and $Y \sim Poisson\left(\frac{x}{\beta}\right)$ for any x, then $P(X \le x) = P(Y \ge \alpha)$.

1.3 Order Statistic

Make sure you know the distribution for $X_{(1)}, X_{(n)}$ and $X_{(j)}$ for $1 < j < n$. *If you want to do something fancy, try to find the joint pdf of any two order statistics. Even more, the joint pdf of given any k order statistics! I am listing only one case here, you can try to prove it if you wish.*

Lemma 1. *Let $X_i, \cdots, X_n \overset{i.i.d}{\sim} f$ with also a distribution function F and assume X_i is continuous. Then given $r < n$, the joint distribution of the order statistics of $X_{(1)}, \cdots, X_{(r)}$ is given by*

$$\mathfrak{f}_r(x_{(1)}, \cdots, x_{(r)}) = \frac{n!}{(n-r)!}\left(\prod_{i=1}^{r} f(x_{(i)}) \cdot \left(1 - F(x_{(r)})\right)^{n-r}\right)$$

1.4 Large Sample Theory

Definitions include: Convergence in Probability, convergence in distribution, CLT, WLLN, Slutsky's Theorem, Continuous Mapping Theorem, (First/Second) order Delta-Method. I am listing some of them:

Theorem 1. *(Delta Methods)*

Let $\{X_n\}_1^\infty$ be a sequence of random variables such that $\mathbb{E}X_n = \mu$, $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} v$ as $n \to \infty$ and g be a real valued function such that $g'(\mu)$ exists and non-zero. Then

$$\sqrt{n}(g(X_n0 - g(\mu)) \xrightarrow{d} g'(\mu) \cdot V.$$

If $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$, and $g'(\mu) = 0, g''(\mu) \neq 0$, then we have

$$n(g(X_n) - g(\mu)) \xrightarrow{d} \frac{1}{2}\sigma^2 g''(\mu)\chi_1^2.$$

1.5 ECDF

Whether you decide to review it or not depends on how you did your midterm Q1.

1.6 Point Estimation

Below, I listed all the important definitions, theorems and some nice results from point estimation. All the exercises in this section are short, and were taken directly from the examples Prof. A. Khalili did in class or from the assignments. Hence no solutions will be provided. Please contact the author if you need the solution to a specific question.

Definition 1. *(Point Estimator and Statistic)*

Given a random sample $X_1, \cdots, X_n \sim f(x, \theta)$, a function $T(\boldsymbol{X}) = T(X_1, \cdots, X_n)$ is an estimator of θ, if it does not depend on the unknown parameter θ, such a $T(\boldsymbol{X})$ is also known as a statistic. The estimator is called an unbiased estimator, if $\mathbb{E}[T(\boldsymbol{X})] = \theta$.

Definition 2. *(Consistency of an Estimator)*

An estimator $T(\boldsymbol{X})$ of θ is consistent, if $T(\boldsymbol{X}) \xrightarrow{P} \theta$.

Exercise 1. *Consider the random sample $X_1, \cdots, X_n \sim Uniform[0, \theta]$, verify the consistency of the estimators $T_1(\boldsymbol{X}) = 2\overline{X}_n, T_2(\boldsymbol{X}) = \frac{n+1}{n}X_{(n)}$. Which one is better? If the random sample is changed to $N(\mu, \sigma^2)$, verify the consistency for $\overline{X}_n$ and S_n^2.*

Exercise 2. *Consider our favourite random sample (apart from $\mathbf{GL}_3(\mathbb{Z}/2\mathbb{Z})$, is $X_1, \cdots, X_n \sim f(x, \theta) = e^{-(x-\theta)}, x \geq \theta$, propose two different consistent estimator of θ, one based on $X_{(1)}$, one based on $\overline{X}_n$. Which one is better?*

In fact we have so many questions for this random sample in exercise 2! Like what is the sufficient statistic? Minimal sufficient statistic? UMVUE? MLE? Method of moment estimator?

Definition 3. *(Regularity Conditions)*

In a random sample $X_1, \cdots, X_n \sim f(x, \theta)$ where θ is one-dimensional, we give the following regularity conditions:

(i) The family $\left\{ f(x, \theta) : \theta \in \Theta \subset \mathbb{R} \right\}$ *has a common support that does not depend on θ.*

(ii) $\dfrac{d}{d\theta} \log f(x, \theta)$ *always exists.*

(iii) For any statistic $h(\boldsymbol{X}) \in L^1$,

$$\frac{d}{d\theta} \int_{\mathcal{S}} h(\boldsymbol{x}) f(\boldsymbol{x}, \theta) dx = \int_{\mathcal{S}} h(\boldsymbol{x}) \frac{d}{d\theta} f(\boldsymbol{x}, \theta) dx$$

Theorem 2. *(Crémer-Rao Lower Bound)*

Under regularly conditions, given a random sample $X_1, \cdots, X_n \sim f(x, \theta)$ with joint pdf denotes as $p_\theta(\boldsymbol{x})$, suppose $T(\boldsymbol{X})$ is an unbiased estimator for $\tau(\theta)$, then

$$Var(T(\boldsymbol{X})) \geq \frac{[\tau'(\theta)]^2}{\mathbb{E}\left\{ \left(\frac{d}{d\theta} \log p_\theta(\boldsymbol{x}) \right)^2 \right\}}$$

The quantity $\mathcal{I}_n = \mathbb{E}\left\{ \left(\frac{d}{d\theta} \log p_\theta(\boldsymbol{x}) \right)^2 \right\}$ *is known as the Fisher information, and $\mathcal{I}_n = n\mathcal{I}_1$.*

> **Exercise 3.** *Given the random sample $X_1, \cdots, X_n \sim Bernoulli(\theta)$, what is the CRLB for all unbiased estimator of θ? Can you find an estimator such that the variance is attained at CRLB?*

Theorem 3. *(Bartlett Identities)*

Under regularity conditions:

First Bartlett Identity is given by

$$\mathbb{E}\left\{ \frac{d}{d\theta} \log p_\theta(\boldsymbol{x}) \right\} = 0, \forall \theta \in \Theta;$$

Second Bartlett Identity is given by

$$\mathbb{E}\left\{ \left(\frac{d}{d\theta} \log p_\theta(\boldsymbol{x}) \right)^2 \right\} = -\mathbb{E}\left\{ \frac{d^2}{d\theta^2} \log p_\theta(\boldsymbol{x}) \right\}.$$

Theorem 4. *(Conditions to attain CRLB)*

Suppose that a random sample $X_1, \cdots, X_n \sim p_\theta(\boldsymbol{x})$, and $T(\boldsymbol{X})$ be an unbiased estimator for $\tau(\theta)$. Then the variance of $T(\boldsymbol{X})$ attains the CRLB iff

$$a(\theta) \cdot \{T(\boldsymbol{X}) - \tau(\theta)\} = \frac{d}{d\theta} \log p_\theta(\boldsymbol{x})$$

for some function $a(\theta)$ for all $\theta \in \Theta$.

Definition 4. *(Exponential Family)*

A random sample $X_1, \cdots, X_n \sim f(x, \theta)$ is said to be of the exponential family, if the joint pdf $p_\theta(\boldsymbol{x})$ takes the form

$$p_\theta(\boldsymbol{x}) = h(\boldsymbol{x}) \cdot c(\theta) \cdot \exp\{\omega(\theta) T(\boldsymbol{X})\}$$

where $h(\boldsymbol{x})$ is a non-negative function of $\boldsymbol{x}$ and $c(\theta)$ a non-negative function of θ, and $T(\boldsymbol{X})$ is a function of $X_1, \cdots, X_n$, $\omega(\theta)$ could be any function of θ. Most importantly, the support $\boldsymbol{x} \in \mathcal{S}$ does not depend on θ. Most most importantly, in this family $\frac{1}{n}\sum T(\boldsymbol{X}_i)$ is the UMVUE of $\tau(\theta) = -\frac{c'(\theta)}{c(\theta)\omega'(\theta)}$.

> **Exercise 4.** *Try to show that the random sample $X_1, \cdots, X_n \sim Poisson(\lambda)$ is of exponential family, and indeed the sample mean $\overline{X}_n$ is the UMVUE of λ.*

Theorem 5. *(Neyman-Fisher Factorization Theorem)*

Let a random sample $X_1, \cdots, X_n \sim p_\theta(\boldsymbol{x})$, a statistics $T(\boldsymbol{X})$ is sufficient iff there exists functions g, h such that

$$p_\theta(\boldsymbol{x}) = g(T(\boldsymbol{x}), \theta) \cdot h(\boldsymbol{x})$$

for all $\theta \in \Theta$ and $\boldsymbol{x} \in \mathcal{X}$. Note that h must be a function that purely depends on $\boldsymbol{x}$, and g depends on $\boldsymbol{x}$ only through $T(\boldsymbol{x})$.

> **Exercise 5.** *Find a sufficient statistic for the given random samples: $X_1, \cdots, X_m \sim Bernoulli(\theta)$, and $Y_1, \cdots, Y_n \sim Uniform[0, \theta]$, and $Z_1, \cdots, Z_l \sim e^{-(x-\theta)}, x \geq \theta$.*
>
> *Have you forgot the indicator function?*
>
> **Exercise 6.** *Now consider the normal family of random samples. If $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$, then what is a sufficient statistic of (μ, σ^2)? If I have two mutually independent random samples $X_1, \cdots, X_m \sim N(\mu_1, \sigma^2)$ and $Y_1, \cdots, Y_n \sim N(\mu_2, \sigma^2)$, then what is a sufficient statistic of (μ_1, μ_2, σ^2)?*

Theorem 6. *(Sufficiency under Transformation)*

If $T(\boldsymbol{X})$ is a sufficient statistics of θ, and we have $T(\boldsymbol{X}) = \phi(T^*(\boldsymbol{X}))$ where ϕ is a measurable function and $T^*(\boldsymbol{X})$ is a statistic, then $T^*(\boldsymbol{X})$ is also sufficient. Or, under any one-to-one transformation g, $g(T(\boldsymbol{X})$ is also sufficient of $g(\theta)$.

> **Exercise 7.** *Convince yourself, that for an exponential family, a sufficient statistic of θ can be $T(\boldsymbol{X}) = \sum X_i$. You will see later, that $T(\boldsymbol{X})$ is also complete!*

Theorem 7. *(Lehmann-Scheffé for Minimum Sufficient Statistic)*

For a parametric family $p_\theta(\cdot)$, suppose a statistic $T(\boldsymbol{X})$ is such that $\forall x, y \in \mathcal{X}$, $T(\boldsymbol{x}) = T(\boldsymbol{y})$ iff $\dfrac{p_\theta(\boldsymbol{x})}{p_\theta(\boldsymbol{y})}$ does not depend on θ, then $T(\boldsymbol{X})$ is a minimum sufficient statistic. In fact any one-to-one function of a minimum sufficient statistics is also minimum sufficient.

Exercise 8. *Again, consider the random samples $X_1, \cdots, X_m \sim Uniform[0, \theta]$ and $Y_1, \cdots, Y_n \sim N(\mu, \sigma^2)$. What are the minimal sufficient statistic for those two?*

Definition 5. *(Completeness of Sufficient Statistic)*

Let X be a random variable with a pdf belonging to a parametric family $\mathcal{F} : \{f_\theta : \theta \in \Theta\}$. This family is said to be complete, if for any measurable function g with $\mathbb{E}[g(x)]$ exists, we have

$$\mathbb{E}[g(x)] = 0, \forall \theta \in \Theta \implies P(g(x) = 0) = 1 \text{ almost surely.}$$

A statistic $T(\boldsymbol{X})$ is complete, if its family of distributions is complete.

Theorem 8. *(Rao-Blackwell)*

Let $U(\boldsymbol{X})$ be an unbiased estimator for $\tau(\theta)$, let $T(\boldsymbol{X})$ be a complete sufficient statistic for the parametric family, then

$$\mathbb{E}\{U(\boldsymbol{X})|T(\boldsymbol{X}) = t\}, \forall t \in \mathcal{S}_T$$

is the UMVUE for $\tau(\theta)$. If the sufficient statistic is not complete, then conditioning on it will only get a better estimator, and it is not necessarily the UMVUE.

Theorem 9. *(Lehmann-Scheffé Uniqueness Theorem)*

Let $T(\boldsymbol{X})$ be a complete sufficient statistic, also let $U(\boldsymbol{X}) = h(T(\boldsymbol{X}))$ for a measurable function h, be an unbiased estimator of $\tau(\theta)$ such that $\mathbb{E}\{U^2(\boldsymbol{X})\} < \infty$. Then $U(\boldsymbol{X})$ is the unique UMVUE.

Exercise 9. *Use Lehmann-Scheffé, find the UMVUE for each parameter of the following random samples:*

$$X_1, \cdots, X_n \sim Bernoulli(\theta), Poisson(\lambda), N(\mu, \sigma^2).$$

Exercise 10. *In the previous exercise, we had two random samples $X_1, \cdots, X_m \sim Bernoulli(\theta)$ and $Y_1, \cdots, Y_n \sim Poisson(\lambda)$. Now find the UMVUE for $\tau(\theta) = \theta(1 - \theta)$, and $\tau(\lambda) = e^{-\lambda}$.*

Theorem 10. *(Relation Between UMVUE and Other Estimators)*

An estimator $U(\boldsymbol{X})$ of $\tau(\theta)$ is the UMVUE iff it is un-correlated with all unbiased estimators of zero, that is,

$$Cov(U(\boldsymbol{X}), \delta(\boldsymbol{X})) = 0$$

for all $\delta(\boldsymbol{X})$ such that $\mathbb{E}\{\delta(\boldsymbol{X})\} = 0$, $\delta(\boldsymbol{X})$ is called the unbiased estimator of zero.

Exercise 11. *(Make sure to look at this! Just in case Khalili really puts this question on the final)*

Let $X \sim Uniform\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right), \theta \in \mathbb{R}$. *Then show that the UMVUE of* $\tau(\theta)$ *must be a constant function.*

Assume $\delta(X)$ is an unbiased estimator of zero, hence $\mathbb{E}[\delta(X)] = 0$. So we have

$$\int_{\theta-1/2}^{\theta+1/2} \delta(X) = 0$$

and by Fundamental Theorem of Calculus we have we have

$$\delta\left(\theta + \frac{1}{2}\right) - \delta\left(\theta - \frac{1}{2}\right) = \frac{d}{d\theta}\int_{\theta-1/2}^{\theta+1/2} \delta(X) = 0$$

Hence we have $\delta(x) = \delta(x+1)$. Now let $U(X)$ be the UMVUE of $\tau(\theta)$, then it is easy to see that $U(X)\delta(X)$ is also an unbiased estimator of zero, and by the previous theorem, we have

$$Cov(U(X), \delta(X)) = \mathbb{E}\{U(X)\delta(X)\} = 0$$

So now we actually have $U(x)\delta(x) = U(x+1)\delta(x+1)$. Hence $U(x) = U(x+1)$, and by definition,

$$\mathbb{E}[U(X)] = \int_{\theta-1/2}^{\theta+1/2} U(X)dx = \tau(\theta)$$

and we use Fundamental Theorem of calculus to get

$$U\left(\theta + \frac{1}{2}\right) - U\left(\theta - \frac{1}{2}\right) = \tau'(\theta)$$

and we know that $\tau'(\theta) = 0$, meaning $\tau(\theta) = C$ where it is a constant! So the UMVUE of $\tau(\theta)$ must be a constant.

Theorem 11. *(Method of Moment Estimator)*

The set up is that, we match the first k *moments of* $f(x, \boldsymbol{\theta})$ *where* $\dim(\boldsymbol{\theta}) = k$ *with the first* k *population mean, if exists. We set them to be equal to set our estimate. i.e we solve the linear system*

$$\begin{aligned}
\frac{X_1 + \cdots + X_n}{n} &\overset{set}{=} \mathbb{E}X \\
\frac{X_1^2 + \cdots + X_n^2}{n} &\overset{set}{=} \mathbb{E}X^2 \\
&\vdots \\
\frac{X_1^k + \cdots + X_n^k}{n} &\overset{set}{=} \mathbb{E}X^k
\end{aligned}$$

Exercise 12. *You better get it, it is so easy. One remark is that is may not always exist, when $\mathbb{E}X = \infty$, like the random sample with pdf $f(x, \theta) = x^{-2}\theta$, where you will see why soon. In a special case where we have a random sample of $Uniform[-\theta, \theta]$, we see that $\mathbb{E}X = 0$ and so we match the second moment.*

Theorem 12. *(Bayesian Estimation)*

Let $\pi(\theta)$ reflect the prior distribution of the unknown parameter θ and $p(\boldsymbol{x}|\theta)$ is the joint pdf of the random sample, then an estimate of θ can be done by

$$\pi(\theta|\boldsymbol{x}) = \frac{p_\theta(\boldsymbol{x}) \cdot \pi(\theta)}{\displaystyle\int_\Theta p_\theta(\boldsymbol{x}) \cdot \pi(\theta) d\theta}$$

where $\pi(\theta|\boldsymbol{x})$ is the posterior distribution. Also under the squared error loss function condition, the Bayes estimate of θ is given by $\mathbb{E}[\pi(\theta|\boldsymbol{x})]$. To find $\pi(\theta|x)$, we may just ignore all the constants as well as the denominator, they are all known as the "normalizing constants", then we may refer to the table to find the distribution.

Exercise 13. *Consider the random sample $X_1, \cdots, X_n \sim Bernoulli(\theta)$, and the prior distribution for θ is given by*

$$\pi(\theta) \sim Beta(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \alpha, \beta > 0$$

Find the Bayes estimator of θ under the squared error loss.

Exercise 14. *Consider the random sample $X_1, \cdots, X_n \sim Uniform(0, \theta)$, and the prior distribution for θ is given by*

$$\pi(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}, \alpha, \beta > 0; \theta > 1$$

Find the Bayes estimator of θ under the squared error loss. You can find the answer to this question in Q9 of the next part.

Theorem 13. *(Invariance of MLE)*

If $\hat{\theta}_n$ is the MLE of θ, then for any function $\tau(\theta)$, $\tau(\hat{\theta}_n)$ is the MLE for $\tau(\theta)$.

Theorem 14. *(Asymptotic Normality of MLE)*

Under all the regularity conditions, if $\hat{\theta}_n$ is the MLE for θ, then

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta)).$$

Exercise 15. *Write a joke such that it contains at least 4 out of the following words: 1 $\mathbf{GL}_3(\mathbb{F}_2)$; 2 MIT; 3 Cows; 4 Electric Cars; 5 Missile; 6 Trivial.*

Finally, here are some distributions, that you might encounter, make sure you know everything about them!!!

- $Bernoulli(\theta)$ **Distribution:**

The pmf is $f(x,\theta) = \theta^x(1-\theta)^{1-x}, 0 < \theta < 1, x\ \{0,1\}$, where θ usually represents the probability of success. The sufficient statistic of θ is $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$, the MLE of θ is $\overline{X}_n$, the Fisher information (for one distribution) is $\mathcal{I}_1(\theta) = \dfrac{1}{\theta(1-\theta)}$, the UMVUE of θ^k is $\binom{n-k}{t-k}/\binom{n}{t}$, where $t = \sum X_i$ is the sufficient statistic!

- $Poisson(\lambda)$ **Distribution:**

The pmf is $f(x,\lambda) = e^{-\lambda}\dfrac{\lambda^x}{x!}, x = 0,1,2,\cdots; \lambda > 0$. The sufficient statistic of θ is $T(\boldsymbol{X}) = \sum X_i$, the MLE of θ is $\overline{X}_n$, the Fisher information (for one distribution) is $\mathcal{I}_1(\lambda) = \dfrac{1}{\lambda}$, the UMVUE of λ^k is $\dfrac{T(T-1)\cdots(T-k+1)}{n^k}$ where $T = \sum X_i$ is the sufficient statistic.

- $Geometric(\theta)$ **Distribution:**

The pmf is $f(x,\theta) = \theta(1-\theta)^{x-1}, x = 1,2,\cdots; 0 < \theta < 1$. The sufficient statistic is $T(\boldsymbol{X}) = \sum X_i$, the MLE is $n/\sum X_i$, the fisher information (for one distribution) is $\mathcal{I}_1(\theta) = \dfrac{1}{\theta^2(1-\theta)}$, and the UMVUE of θ is $\binom{t-2}{n-2}/\binom{t-1}{n-1}$ where $t = \sum X_i$ is the sufficient statistic.

- $Uniform(0,\theta)$ **Distribution:**

The pdf is $f(x,\theta) = \dfrac{1}{\theta}\cdot\chi\{0 < X < \theta\}$, a sufficient statistic is $X_{(n)}$, the MLE is $X_{(n)}$, the UMVUE of θ^k is $\dfrac{n+k}{n}X_{(n)}^k$.

- **Shifted Exponential Distribution:**

The pdf is $f(x,\theta) = e^{-(x-\theta)}\cdot\chi\{x > \theta\}$, in this family we have $\mathbb{E}X = \theta + 1$ and $Var(X) = 1$, and a sufficient statistic is $X_{(1)}$, the cdf of $X_{(1)}$ is $F(x) = 1 - e^{n(\theta-x)}$ and the pdf of $X_{(1)}$ is $f(x) = ne^{n(\theta-x)}$, the UMVUE of θ is $X_{(1)} - \frac{1}{n}$.

- **Exponential Distribution:**

The pdf is $f(x,\theta) = \theta e^{-\theta x}, x > 0$, in this family we have $\mathbb{E}X = \frac{1}{\theta}$, and $Var(X) = \frac{1}{\theta^2}$, a sufficient statistic is $T(\boldsymbol{X}) = \sum X_i$, the MLE is $\frac{n}{\sum X_i}$, the Fisher information (for one distribution) is $\mathcal{I}_1(\theta) = \frac{1}{\theta^2}$, the UMVUE of θ is $\frac{n-1}{\sum X_i}$. In this family, a well known integral is important:

$$\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx = \Gamma(\alpha)\beta^\alpha$$

Furthermore, you can see the UMVUE of $\theta^k, k < n$ is

$$\frac{\Gamma(n)}{\Gamma(n-k)} \cdot \frac{1}{n^k} \cdot \frac{n}{\sum X_i}.$$

1.7 Confidence Interval

Definition 6. *(Interval Estimator)*

Let $L(\boldsymbol{X}), U(\boldsymbol{X})$ be two statistics such that $L(\boldsymbol{x}) \leq U(\boldsymbol{x}), \forall x \in \mathcal{X}$. A random interval $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ is called an interval estimator or confidence interval with confidence level $1-\alpha, \alpha \in (0,1)$ if $P(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})) = 1-\alpha$.

Note that it is **wrong** to say that $(L(\boldsymbol{x}), U(\boldsymbol{x}))$ (post-experimental data) captures θ with probability $1-\alpha$. The interpretation is that *this interval estimator either includes θ or not, basically it captures θ with probability 0 or 1. if we were to repeat the experiment and compute similar confidence intervals for θ, we expect that $100(1-\alpha)\%$ of those post-experimental intervals to capture θ.*

Definition 7. *(Pivot Quantity)*

A random function $Q(\boldsymbol{x}, \theta)$ is called a pivot quantity (PQ) iff its distribution does not depend on the parameter θ, and Q is a function of $\boldsymbol{X}$ and θ only.

Note that the function Q might include θ, but its overall distribution $Q(\boldsymbol{x}, \theta)$ is free of any parameters! For example if $X \sim N(0, \theta^2)$, then $Q(x, \theta) = \frac{1}{\theta} X \sim N(0,1)$, which is free of any parameter and it is a known distribution!

Once we have a PQ, and the confidence level $1-\alpha$ is given, we may find constants c_1, c_2 such that

$$P(c_1 \leq Q(\boldsymbol{x}, \theta) \leq c_2) = 1-\alpha,$$

then having c_1, c_2 we can solve in terms of θ to get

$$P(L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})) = 1-\alpha.$$

Mostly, the PQ is chosen based on a sufficient statistic.
Below are common random samples and their corresponding PQ and confidence intervals:

(i) **Confidence Interval for μ in a normal family**:

- If σ^2 is known, then

$$Q(\boldsymbol{X}, \mu) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0,1)$$

and the $100(1-\alpha)\%$ C.I is given by

$$\left(\overline{X}_n - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

• If σ^2 is unknown, then

$$Q(\boldsymbol{X}, \mu) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t(n-1)$$

and the $100(1-\alpha)\%$ C.I is given by

$$\left(\overline{X}_n - t(n-1, \alpha/2) \cdot \frac{S_n}{\sqrt{n}}, \overline{X}_n + t(n-1, \alpha/2) \cdot \frac{S_n}{\sqrt{n}}\right)$$

Note: z_p is called the p-th quantile in a standard normal distribution, and $P(Z < z_p) = p$, where $Z \sim N(0,1)$. The same idea applies in $t(n-1, \alpha/2)$, where $n-1$ indicates the degrees of freedom, and $\alpha/2$ is the quantle.

Exercise 16. *A manufacturer developed a new gunpowder and tested it in eight shells. The resulting muzzle velocities, in feet per second, were:*

$$3005, 2925, 2935, 2965, 2995, 3005, 2937, 2905.$$

Assume that the velocities are iid sample from a $N(\mu, \sigma^2)$. Compute a 95% confidence interval for μ. Provide interpretation for your interval.

So in this example, we see that σ^2 is unknown. So we replace it by S_n. Note that the S_n of this sample can be calculated by

$$S_n = \sqrt{S_n^2} = \sqrt{\frac{1}{7}\sum_{i=1}^{8}(X_i - \overline{X}_n)^2} \approx 39.09$$

From the C.I above, we construct

$$\left(2959 - t(7, 0.025) \cdot \frac{39.09}{\sqrt{8}}, 2959 + t(7, 0.025) \cdot \frac{39.09}{\sqrt{8}}\right).$$

From the t-table, $t(7, 0.025) = 2.365$, so for the 95% confidence interval, our final answer is $(2926.38, 2991.62)$. The interpretation is that, *this interval estimator either includes θ or not, basically it captures θ with probability* 0 *or* 1. *if we were to repeat the experiment and compute similar confidence intervals for μ, we expect that* $100(1-\alpha)\%$ *of those post-experimental intervals to capture μ.*

(ii) **Confidence interval for σ^2 in a normal family:**

• If μ is unknown, then

$$Q(\boldsymbol{X}, \sigma^2) = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

and the $100(1-\alpha)\%$ C.I is given by

$$\left(\frac{(n-1)S_n^2}{\chi^2_{(n-1,\alpha/2)}}, \frac{(n-1)S_n^2}{\chi^2_{(n-1,1-\alpha/2)}}\right).$$

• if μ is known, then

$$Q(\boldsymbol{X}, \sigma^2) = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$$

and the $100(1-\alpha)\%$ C.I is given by

$$\left(\frac{(n-1)\sum(X_i-\mu)^2}{\chi^2_{(n,\alpha/2)}}, \frac{(n-1)\sum(X_i-\mu)^2}{\chi^2_{(n,1-\alpha/2)}} \right)$$

Exercise 17. *Assume that the number of days needed to hatch an egg of a certain type of a rare lizard is distributed Normally. Using incubator,* 13 *eggs from different nests separately hatched. The sample mean is* 18.97 *weeks and the sample standard deviation is* $\sqrt{10.7}$ *weeks. Find a* 90% *confidence interval for the population variance. Provide interpretation for the interval.*

Here, it is the case that both μ, σ^2 are unknown, so we construct the C.I based on

$$\left(\frac{(n-1)S_n^2}{\chi^2_{(n-1,\alpha/2)}}, \frac{(n-1)S_n^2}{\chi^2_{(n-1,1-\alpha/2)}} \right) = \left(\frac{12 \cdot 10.7}{\chi^2_{(12,0.05)}}, \frac{12 \cdot 10.7}{\chi^2_{(12,0.95)}} \right) = (6.107, 24.569).$$

Note that in this example sample mean is not population mean μ! It is there to confuse you :) The interpretation is that, *this interval estimator either includes θ or not, basically it captures θ with probability* 0 *or* 1. *if we were to repeat the experiment and compute similar confidence intervals for μ, we expect that* $100(1-\alpha)\%$ *of those post-experimental intervals to capture μ, which is the same as above.*

(iii) **Confidence Interval Approximation for μ of non-normal large sample:**

We will assume that in a large random sample (*usually $n > 25$ is good enough*) we have $X_1, \cdots, X_n \sim f$, $\mathbb{E}X = \mu$, and $VarX = \sigma^2$, $EX^4 < \infty$, and we consider the asymptotic approximate C.I for μ, as $n \to \infty$.

• If σ is known, then we have

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

so the $100(1-\alpha)\%$ C.I approximate is

$$\left(\overline{X}_n - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right).$$

• If σ is unknown, then we have

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$$

so the $100(1-\alpha)\%$ C.I approximate is

$$\left(\overline{X}_n - z_{\alpha/2} \cdot \frac{S_n}{\sqrt{n}}, \overline{X}_n + z_{\alpha/2} \cdot \frac{S_n}{\sqrt{n}} \right).$$

Exercise 18. *Shopping times of* 64 *randomly selected customers in a supermarket averaged* 33 *minutes with a standard deviation of* 16 *minutes. Construct an approximate* 90% *confidence interval for the mean shopping time per customer. Provide interpretation for the interval.*

Standard deviation is S_n, not the square!!!

Here, if we use large sample theory and the central limit theorem, we will have

$$\frac{\sqrt{64}(33-\mu)}{16} \xrightarrow{d} N(0,1)$$

Here $\alpha = 0.1$, so we look at the normal table and see that $z_{\alpha/2} = z_{0.05} \approx 1.64$, hence we have the C.I

$$\left(33 - 1.64 \times \frac{16}{8}, 33 + 1.64 \times \frac{15}{8}\right) = (29.72, 36.28).$$

Remark: *Since we are interested in $z_{0.05}$. if we can not find the exact value, i.e we know $z_{0.0495} = 1.65$ and $z_{0.505=1.64}$, we then may take the average to get a better approximate of $z_{0.05} \approx$* 1.645.

(iv) **Mean Difference in Two Mutually Independent Normal Random Samples:**

Given two mutually independent random samples $X_1, \cdots, X_m \sim N(\mu_1, \sigma^2)$ and $Y_1, \cdots, Y_n \sim N(\mu_2, \sigma^2)$, we are interested in constructing the confidence interval for $\mu_1 - \mu_2$. The PQ is

$$Q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\mu}) = \frac{(\overline{X}_m - \overline{Y}_n) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$$

where

$$S^2 = \frac{1}{m+n-2}\left\{\sum_{i=1}^{m}(X_i - \overline{X}_m)^2 + \sum_{j=1}^{n}(Y_i - \overline{Y}_n)^2\right\}$$

so the $100(1-\alpha)\%$ C.I is given by

$$\left((\overline{X}_m - \overline{Y}_n) - t(m+n-2, \alpha/2) \cdot S\sqrt{\frac{1}{m} + \frac{1}{n}}, (\overline{X}_m - \overline{Y}_n) + t(n+m-2, \alpha/2) \cdot S\sqrt{\frac{1}{m} + \frac{1}{n}}\right)$$

Exercise 19. *In a packing plant, a machine packs cartons with jars. It is supposed that a new machine will pack faster on the average than the machine currently used. To test that hypothesis, the times it takes each machine to pack ten cartons are recorded. The results in seconds are:*

$$\textit{old}: 42.7, 43.8, 42.5, 43.1, 44.0, 43.6, 43.3, 43.5, 41.7, 44.1;$$
$$\textit{new}: 42.1, 41.3, 42.4, 43.2, 41.8, 41.0, 41.8, 42.8, 42.3, 42.7.$$

Construct a 95% *confidence interval for the difference in the respective means. Provide interpretation for the interval. (Assume that the timings for the old and new machines are independent i.i.d samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively, and $\sigma_1 = \sigma_2$.*

Here, denote X_i to be the old sample, and Y_j to be the new sample. Then from the data provided we can compute that $\overline{X} = 43.23$ and $\overline{Y} = 42.14$, so $\overline{X} - \overline{Y} = 1.09$, and

$$S^2 = \frac{1}{18}\left\{\sum_{i=1}^{10}(\overline{X}_i - 43.23)^2 + \sum_{j=1}^{10}(\overline{Y}_j - 42.14)^2\right\}$$

(v) **Mean Difference in Two Mutually Independent Non-normal Random Samples:**

This set up is roughly the same as the previous one, but here we removed the normal restriction, and we use central limit theorem to get the approximate of C.I for large n, m. We know that the PQ is

$$Q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}) = \frac{(\overline{X}_m - \overline{Y}_n) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \xrightarrow{d} N(0,1).$$

and the $100(1-\alpha)\%$ C.I is given by

$$\left(\overline{X}_m - \overline{Y}_n - z_{\alpha/2} \cdot \sqrt{\frac{S_m^2}{m} + \frac{S_n^2}{n}}, \overline{X}_m - \overline{Y}_n + z_{\alpha/2} \cdot \sqrt{\frac{S_m^2}{m} + \frac{S_n^2}{n}}\right)$$

Note that as $n \to \infty$, $S_m^2 \xrightarrow{P} \sigma_1^2$ and $S_n^2 \xrightarrow{P} \sigma_2^2$, so we can swap to whichever is easier for us to compute.

Exercise 20. *We wish to compare the daily intake of selenium in two regions. In each region,* 30 *adults were tested and the results (in mg/day) were:* $\overline{x}_n = 167.1, s_n = 24.3, \overline{y}_m = 140.9, s_m = 17.6$ *Find a* 95% *approximate confidence interval for the difference in mean daily intake of selenium in the two regions. Provide interpretation for the interval.*

This is the case where we will find the C.I for the difference of mean in two non-normal random samples. So we apply the above formula, we know that $\overline{X}_m - \overline{Y}_n = 26.2$, $\sqrt{\frac{s_m^2}{m} + \frac{s_n^2}{n}} \approx 5.477986$ and here $\alpha = 0.05$, so $z_{\alpha/2} = z_{0.025}$, and hence we have the 95% C.I to be $26.2 \pm z_{0.025} \cdot 5.477986$. From the normal table, $z_{0.025} = 1.96$, so finally we have $(15.46315, 36.93685)$.

Remark: *Please make sure how to read a normal table!! Here you want to find $z_{\alpha/2}$, where by definition we have $P(X < z_{\alpha/2}) = \alpha/2$, so you will go through in the middle of the table to get $\alpha/2$ first (in our table would be $1 - \alpha/2$ and then see the coordinates correspond to z! That will give you the correct x value! Make sure you know it! Make sure you know it! Don't wait until you are in the gym and realized you're an idiot.*

(vi) **Population Proportion:**

Now consider we have a random sample $X_1, \cdots, X_n \sim Bernoulli(\theta)$, then we know that by central limit theorem, a PQ is

$$Q(\boldsymbol{X}, \theta) = \frac{\overline{X}_n - \theta}{\sqrt{\theta(1-\theta)/n}} \xrightarrow{d} N(0,1)$$

also for large sample, we have $\overline{X}_n \xrightarrow{P} \theta$, denote $\hat{\theta}_n = \overline{X}_n$, we also have an alternative PQ

$$Q(\boldsymbol{X}, \theta) = \frac{\hat{\theta}_n - \theta}{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}} \xrightarrow{d} N(0,1).$$

The advantage for this PQ is that we can easily separate θ and get the $100(1-\alpha)\%$ C.I:

$$\left(\hat{\theta}_n - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}, \hat{\theta}_n + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}_n(1-\hat{\theta}_n)}{n}}\right)$$

Exercise 21. *A sample of $n = 1000$ voters, randomly selected from a city, showed 560 in favour of candidate Jones. Find an approximate 99% confidence interval for the population proportion in favour of candidate Jones. Provide interpretation for the interval.*

Here we can easily see that in this Bernoulli random sample, we have $\hat{\theta}_n = \overline{X}_n = 0.56$, hence we construct the C.I by

$$\left(0.56 - z_{0.005} \cdot \sqrt{\frac{0.56 \cdot 0.44}{1000}}, 0.56 + z_{0.005} \cdot \sqrt{\frac{0.56 \cdot 0.44}{1000}}\right) = 0.56 \pm z_{0.005} \cdot 0.0157$$

From the normal table, we see that $z_{0.005} \approx 2.575$.

(vii) **Difference in Two Population Proportion:**

Here, we have two mutually independent random samples $X_1, \cdots, X_m \sim \mathit{Bernoulli}(\theta_1)$ and $Y_1, \cdots, Y_n \sim \mathit{Bernoulli}(\theta_2)$. We are interested in the C.I of $\theta_1 - \theta_2$. Similarly, denote $\hat{\theta}_1 = \overline{X}_m, \hat{\theta}_2 = \overline{Y}_n$, and the improved PQ is

$$Q(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}) = \frac{(\hat{\theta}_1 - \theta_1) - (\hat{\theta}_2 - \theta_2)}{\sqrt{\hat{\theta}_1(1-\hat{\theta}_1)/m + \hat{\theta}_2(1-\hat{\theta}_2)/n}} \xrightarrow{d} N(0,1).$$

and the $100(1-\alpha)\%$ C.I is now

$$\left((\hat{\theta}_1 - \hat{\theta}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{m} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n}}, (\hat{\theta}_1 - \hat{\theta}_2) + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{m} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n}}\right)$$

Exercise 22. *A medical researcher conjectures that smoking can result in wrinkled skin around the eyes. The researcher recruited 150 smokers and 250 nonsmokers to take part in an observational study and found that 95 of the smokers and 105 of the nonsmokers were seen to have prominent wrinkles around the eyes (based on a standardized wrinkle score administered by a person who did not know if the subject smoked or not). Find an approximate 95% confidence interval for the difference in the proportions of people who have wrinkled skin around their eyes in the two populations. Provide interpretation for the interval.*

Here, we have two independent random samples, let $X_1, \cdots, X_{150} \sim Bernoulli(\theta_1)$ to be the smokers sample and $Y_1, \cdots, Y_{250} \sim Bernoulli(\theta_2)$ to be the nonsmokers sample, where θ_1, θ_2 denotes the proportion of populations who have wrinkled skin, thus $\hat{\theta}_1 = 95/150 = 0.633$ and $\hat{\theta}_2 = 105/250 = 0.42$. Also in this case we have $\alpha = 0.05$ hence $\alpha/2 = 0.025$ and from the normal table we have $z_{\alpha/2} = z_{0.025} = 1.96$, also $\hat{\theta}_1 - \hat{\theta}_2 = 0.213$, $\sqrt{\dfrac{\hat{\theta}_1(1-\hat{\theta}_1)}{m} + \dfrac{\hat{\theta}_2(1-\hat{\theta}_2)}{n}} \approx 0.0502$, and thus our 95% C.I is $(0.1146, 0.3114)$. *As for interpretation? You know that.*

(viii) **Approximate Using MLE Theory:**

Recall that, let a random sample $X_1, \cdots, X_n \sim f(x, \theta)$ where θ is a one-dimensional parameter, and $\hat{\theta}_n$ to be the MLE of θ, then we know that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \mathcal{I}_1^{-1}(\hat{\theta}_n)\right)$$

One remark is that since we have convergence in probability $\hat{\theta}_n \xrightarrow{P} \theta$, so we can replaced by $\hat{\theta}_n$ in Fisher information, which would be easier to separate θ in the calculation.

So we have the $100(1-\alpha)\%$ given by

$$\left(\hat{\theta}_n - z_{\alpha/2} \cdot \sqrt{\frac{1}{n}[\mathcal{I}_1(\hat{\theta}_n)]^{-1}}, \hat{\theta}_n + z_{\alpha/2} \cdot \sqrt{\frac{1}{n}[\mathcal{I}(\hat{\theta}_n)]^{-1}}\right).$$

Also we may use the "empirical Fisher" to get our estimate of $\mathcal{I}_1(\hat{\theta})$:

$$\mathcal{I}_1(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\log f(x_i, \theta)\bigg|_{\theta=\hat{\theta}_{MLE}}\right)^2 \overset{\text{under regularity conditions}}{=} -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\theta^2}\log f(x_i, \theta)\bigg|_{\theta=\hat{\theta}_{MLE}}\right)$$

Recall the delta-method and the invariance of MLE. If $\hat{\theta}_n$ is the MLE of θ, and then for any function g, $g(\hat{\theta}_n)$ is the MLE of $g(\theta)$, and the first order delta-method shows us that if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\hat{\theta}_n))$, then

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2 \cdot \mathcal{I}_1^{-1}(\hat{\theta}_n)).$$

and the $100(1-\alpha)\%$ C.I approximate is

$$\left(g(\hat{\theta}_n) - z_{\alpha/2} \cdot \sqrt{\frac{1}{n}\mathcal{I}_1^{-1}(\hat{\theta}_n) \cdot |g'(\hat{\theta}_n)|^2}\ ,\ g(\hat{\theta}_n) + z_{\alpha/2} \cdot \sqrt{\frac{1}{n}\mathcal{I}_1^{-1}(\hat{\theta}_n) \cdot |g'(\hat{\theta}_n)|^2}\right)$$

Exercise 23. *Suppose a random sample $X_1, \cdots, X_n \sim Poisson(\lambda)$ and λ is the unknown parameter. Using the MLE theory, construct a $100(1-\alpha)\%$ approximate two-sided confidence interval for λ. Then find the $100(1-\alpha)\%$ C.I for λ^2, and $e^{-\lambda}$.*

Here, we can easily see that the MLE of a Poisson random sample is just the sample mean, $\hat{\theta}_n = \overline{X}_n$. Now we find the Fisher information. The log-pdf of one sample is

$$\log f(X = k, \lambda) = \log e^{-\lambda}\frac{\lambda^k}{k!} = -\lambda + k\log\lambda - \log k!.$$

The second partial derivative is

$$\frac{\partial^2 \log f(X=k,\lambda)}{\partial \lambda^2} = -k\frac{1}{\lambda^2}$$

and hence we have

$$\mathcal{I}_1(\lambda) = -\mathbb{E}\left\{-X\frac{1}{\lambda^2}\right\} = \frac{1}{\lambda}, \mathcal{I}_1^{-1}(\lambda) = \lambda$$

So the MLE theory says that

$$\sqrt{n}\left(\overline{X}_n - \lambda\right) \xrightarrow{d} N\left(0, \mathcal{I}_1^{-1}(\hat{\theta}_n)\right)$$

and given that $\mathcal{I}_1^{-1}(\hat{\theta}) = \overline{X}_n$, hence the $100(1-\alpha)\%$ C.I is

$$\left(\overline{X}_n - z_{\alpha/2}\cdot\sqrt{\frac{1}{n}\cdot\overline{X}_n}\ ,\ \overline{X}_n + z_{\alpha/2}\cdot\sqrt{\frac{1}{n}\cdot\overline{X}_n}\right).$$

For λ^2 and $e^{-\lambda}$, we will then apply delta-method to find the C.I, we define $g(\lambda) = \lambda^2$, and $h(\lambda) = e^{-\lambda}$, hence it is easy to see that $\hat{\lambda}^2_{MLE} = \overline{X}_n^2$ and $e^{-\lambda}_{MLE} = e^{-\overline{X}_n}$. Then delta-method says that

$$\sqrt{n}(\overline{X}_n^2 - \lambda^2) \xrightarrow{d} N\left(0, 4\lambda^2\cdot\overline{X}_n\right)$$

which is,

$$\frac{\sqrt{n}(\overline{X}_n^2 - \lambda^2)}{2\lambda\sqrt{\overline{X}_n}} \xrightarrow{d} N(0,1)$$

use our large sample theory, we have $\overline{X}_n \xrightarrow{P} \lambda$, so we replace the λ in the denominator by $\overline{X}_n$, and we get the C.I as

$$\left(\overline{X}_n^2 - z_{\alpha/2}\cdot 2\overline{X}_n\sqrt{\frac{\overline{X}_n}{n}}, \overline{X}_n^2 + z_{\alpha/2}\cdot 2\overline{X}_n\sqrt{\frac{\overline{X}_n}{n}}\right).$$

Similarly, in $e^{-\lambda}$, we have

$$\frac{\sqrt{n}(e^{-\overline{X}_n} - e^{-\lambda})}{\lambda e^{-\lambda}\cdot\sqrt{e^{-\overline{X}_n}}} \xrightarrow{d} N(0,1)$$

and the C.I is

$$\left(e^{-\overline{X}_n} - z_{\alpha/2}\cdot 2e^{-\overline{X}_n}\cdot e^{-e^{\overline{X}_n}}\cdot\sqrt{\frac{e^{-\overline{X}_n}}{n}}\ ,\ e^{-\overline{X}_n} + z_{\alpha/2}\cdot 2e^{-\overline{X}_n}\cdot e^{-e^{\overline{X}_n}}\cdot\sqrt{\frac{e^{-\overline{X}_n}}{n}}\right).$$

(ix) **One-Sided Confidence Interval:**

Here we use an example to indicate how to deal with one sided confidence interval. The idea is easy, in this case we don't do $\alpha/2$, but use α directly, but need to verify which side we need to use.

1.8 Hypothesis Testing

Definition 8. *(Hypothesis Test)*

A statistical hypothesis test is a decision rule that uses the data to infer which two mutually exclusive hypothesis, that reflect two completing hypothetical states of the nature is correct. The decision rule partitions the sample space $\mathcal{X}$ into 2 disjoint regions that respectively reflect support for the null hypothesis $\mathcal{H}_0$ and the alternative hypothesis $\mathcal{H}_1$, where $\mathcal{X} = \mathcal{H}_0 \bigcup \mathcal{H}_1$ and $\mathcal{H}_0 \bigcap \mathcal{H}_1 = \varnothing$. Our goal would be to use the data to decide whether the parameter of interest θ is whether $\theta \in \mathcal{H}_0$ or $\theta \in \mathcal{H}_1$. Unlike point estimation and confidence interval, we do not perform any estimate on θ.

Definition 9. *(Test Function)*

Suppose a test $\mathcal{H}_0$ and $\mathcal{H}_1$ partitioning the sample space $\mathcal{X}$ into two disjoint regions $\mathcal{R}$ and $\mathcal{R}^C$, and we will reject $\mathcal{H}_0$ if $\boldsymbol{x} \in \mathcal{R}$, such $\mathcal{R}$ is called the critical region. We may formulate the test as a function:

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \textit{if } \boldsymbol{x} \in \mathcal{R}, \textit{ and we reject } \mathcal{H}_0 \\ 0 & \textit{if } \boldsymbol{x} \in \mathcal{R}^C, \textit{ and we do not reject } \mathcal{H}_0 \end{cases}$$

Definition 10. *(Test Errors)*

There are two types of errors: Type one error is when $\mathcal{H}_0$ is rejected when $\mathcal{H}_0$ is indeed true; Type two error is $\mathcal{H}_0$ is not rejected when $\mathcal{H}_1$ is indeed true.

Theorem 15. *(Neyman-Pearson)*

Let $\phi(\boldsymbol{X}) = \begin{cases} 1 & \textit{if } p(\boldsymbol{x}, \theta_1) > kp(x, \theta_0) \\ \\ 0 & \textit{if } p(\boldsymbol{x}, \theta_1) < kp(\boldsymbol{x}, \theta_0) \end{cases}$, and k is such that $P(\textit{rejecting } \mathcal{H}_0) = \alpha$, (we reject $\mathcal{H}_0$ if $\phi(\boldsymbol{X}) = 1$)then ϕ is the UMP test in the class of all tests ϕ^* with the same level α, α is called the significance level. Hence the UMP test has a rejection region

$$\mathcal{R} := \left\{ \boldsymbol{x} \in \mathcal{X}, \frac{p(\boldsymbol{x}, \mathcal{H}_1)}{p(\boldsymbol{x}, \mathcal{H}_0)} > k \right\}, \ k \textit{ is chosen such that } P(\boldsymbol{x} \in \mathcal{R}) \leq \alpha.$$

> **Exercise 24.** *Suppose we have a random sample $X_1, \cdots, X_n \sim N(\mu, 1)$, and $\mu = \{0, 1\}$. So we would like to test that $\mathcal{H}_0 : \mu = 0$ and $\mathcal{H}_1 : \mu = 1$.*

To use NP lemma, we first construct the ratio:

$$\frac{p(\boldsymbol{x}, \mu = 1)}{p(\boldsymbol{x}, \mu = 0)} = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum(x_i - 1)^2\right\}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum(x_i)^2\right\}} = e^{-\frac{1}{2}\sum(x_i-1)^2 + \frac{1}{2}\sum x_i^2} = e^{n\overline{X}_n - n/2}$$

Then the NP lemma says that we will reject $\mathcal{H}_0$ if $\dfrac{p(\boldsymbol{x}, \mu = 1)}{p(\boldsymbol{x}, \mu = 0)} > k$ for some k, so we solve for $e^{n\overline{X}_n - n/2} > k$, and we get $\overline{x}_n > k^* = \frac{\ln k}{n} + \frac{1}{2}$, and this is the rejection region. Now given significance level α, the type one error is given by

$$P(\text{Reject } \mathcal{H}_0 \text{ when it is true}) = P(\overline{x}_n > k \big| \mu = 0) = \alpha.$$

Then we use the fact that $\overline{X}_n \sim N\left(0, \frac{1}{n}\right)$ (the case when $\mathcal{H}_0$ is true), and we have

$$P\left(\frac{\sqrt{n}(\overline{X}_n - 0)}{1} > \sqrt{n}k^*\right) = \alpha$$

and then we can refer to the normal table to solve for k^*. The same idea applies for type two error, where in this case we have

$$P(\text{Not rejecting } \mathcal{H}_0 \text{ when it is false}) = P(\overline{x}_n < k^* \big| \mu = 1).$$

Theorem 16. *(Likelihood Ratio Test)*

Consider the random sample $X_1, \cdots, X_n \sim f(x, \theta)$, and we have $\mathcal{H}_0$ and $\mathcal{H}_1$ as two tests, and we define the likelihood ratio (LR) statistic to be

$$\lambda_n(\boldsymbol{X}) = \frac{L_n(\hat{\theta}_{MLE,\mathcal{H}_0})}{L_n(\hat{\theta}_{MLE,\Theta})}$$

A test based on LR statistic has the following form

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & \lambda_n(\boldsymbol{X}) < C \text{ (reject } \mathcal{H}_0) \\ 0 & \lambda_n(\boldsymbol{X}) > C \end{cases}$$

for $C \in [0, 1]$ and the rejection region takes the form $\mathcal{R} := \{\boldsymbol{x} \in \mathcal{X} : \lambda_n(\boldsymbol{X}) < C\}$.

Theorem 17. *(Asymptotic Approximation of LR)*

At significance level α, the rejection region ($\lambda_n(\boldsymbol{X}) < C$) of the LR-based test under regularity conditions for large n is approximately

$$\mathcal{R} := \Big\{\boldsymbol{x} \in \mathcal{X} : -2\log[\lambda_n(\boldsymbol{X})] \geq \chi^2_{d,\alpha}\Big\}, d = \dim\Theta - \dim\Theta_0$$

where

$$-2\log[\lambda_n(\boldsymbol{X})] = 2\Big\{\sup_{\theta\in\Theta}\ell_n(\theta) - \sup_{\theta\in\Theta_0}\ell_n(\theta)\Big\}$$

Also, we have a general formula to solve for hypothesis testing. We list some cases here:

(i) **Testing the Mean in a (Asymptotically) Normal Sample with Known Variance**

Suppose we know the random sample takes the form $N(\mu, \sigma^2)$ where σ^2 is known, and we wish to test:

$$\mathcal{H}_0 : \mu = \mu_0, \mathcal{H}_1 : \mu \neq \mu_0 (\mu > \mu_0)(\mu < \mu_0)$$

and we first compute the value under the null hypothesis:

$$z = \frac{\sqrt{n}(\overline{x}_n - \mu_0)}{\sigma} \text{ In a large sample we may replace } (\sigma \text{ by } s_n)$$

and we will reject $\mathcal{H}_0$ at significance level α if:

$$|z| > z_{\alpha/2}(z > z_\alpha)(z < -z_\alpha).$$

(ii) **Testing the Mean in a Normal Sample with Unknown Variance**

Suppose we know the random sample takes the form $N(\mu, \sigma^2)$ where σ^2 is unknown, and we wish to test:

$$\mathcal{H}_0 : \mu = \mu_0, \mathcal{H}_1 : \mu \neq \mu_0(\mu > \mu_0)(\mu < \mu_0)$$

and we first compute the value under the null hypothesis:

$$t = \frac{\sqrt{n}(\overline{x}_n - \mu_0)}{s_n}$$

and we will reject $\mathcal{H}_0$ at significance level α if:

$$|t| > t_{\alpha/2,n-1}(t > t_{\alpha,n-1})(t < -t_{\alpha,n-1}).$$

(iii) **Testing the Variance in a Normal Sample with Unknown Mean and Variance**

Suppose we know the random sample takes the form $N(\mu, \sigma^2)$ where μ, σ^2 is unknown, and we wish to test:

$$\mathcal{H}_0 : \sigma^2 = \sigma_0^2, \mathcal{H}_1 : \sigma^2 \neq \sigma_0^2(\sigma^2 > \sigma_0^2)(\sigma^2 < \sigma_0^2)$$

and we first compute the value under the null hypothesis:

$$\chi = \frac{(n-1)s^2}{\sigma_0^2}$$

and we will reject $\mathcal{H}_0$ at significance level α if:

$$\chi^2 > \chi^2_{\alpha/2,n-1}\text{or} < \chi^2_{1-\alpha/2,n-1}(\chi^2 > \chi^2_{\alpha,n-1})(\chi^2 < \chi^2_{1-\alpha,n-1})$$

(iv) **Testing the Ratio in a Bernoulli Sample**

Suppose we know the random sample takes the form $Bernoulli(\theta)$ where θ is unknown, and we wish to test:

$$\mathcal{H}_0 : \theta = \theta_0, \mathcal{H}_1 : \theta \neq \theta_0(\theta > \theta_0)(\theta < \theta_0)$$

and we first compute the value under the null hypothesis:

$$z = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$$

and we will reject $\mathcal{H}_0$ at significance level α if:

$$|z| > z_{\alpha/2}(z > z_\alpha)(z < -z_\alpha).$$

(v) **Testing the Difference in Mean of Two Bernoulli Samples**

Suppose we have two mutually independent large random samples ($\geq$ 25) $X_1, \cdots, X_m \sim Bernoulli(\theta_1)$ and $Y_1, \cdots, Y_n \sim Bernoulli(\theta_2)$, and we wish to test

$$\mathcal{H}_0 : \theta_1 - \theta_2 = D_0; \mathcal{H}_1 : \theta_1 - \theta_2 \neq D_0 (\theta_1 - \theta_2 > D_0)(\theta_1 - \theta_2 < D_0)$$

and we first compute the value under the null hypothesis:

If $D_0 = 0$, then

$$z = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{m} + \frac{\hat{\theta}(1-\hat{\theta})}{n}}}, \hat{\theta} = \frac{x+y}{m+n}, \text{where } x/m = \hat{\theta}_1, y/n = \hat{\theta}_2$$

If $D_0 \neq 0$, then

$$z = \frac{\hat{\theta}_1 - \hat{\theta}_2 - D_0}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{m} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n}}}$$

and we will reject $\mathcal{H}_0$ at significance level α if

$$|z| > z_{\alpha/2} (z > z_\alpha)(z < z_\alpha).$$

(vi) **Testing the Difference in Mean of Two Normal Samples with Known Variance**

Suppose we have two mutually independent large samples ($>$ 25) $X_1, \cdots, X_m \sim N(\mu_1, \sigma_1^2)$ and $Y_1, \cdots, Y_n \sim N(\mu, \sigma_2^2)$ where σ_1, σ_2 are known. We wish to test

$$\mathcal{H}_0 : \mu_1 - \mu_2 = D_0; \mathcal{H}_1 : \mu_1 - \mu_2 \neq D_0 (\mu_1 - \mu_2 > D_0)(\mu_1 - \mu_2 < D_0)$$

We first compute the value under the null hypothesis:

$$z = \frac{\overline{x}_m - \overline{y}_n - D_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

and we will reject $\mathcal{H}_0$ at significance level α if

$$|z| > z_{\alpha/2} (z > z_\alpha)(z < -z_\alpha)$$

(vii) **Testing the Difference in Mean of Two Normal Samples with Unknown Variance**

Suppose we have two mutually independent large samples ($>$ 25) $X_1, \cdots, X_m \sim N(\mu_1, \sigma_1^2)$ and $Y_1, \cdots, Y_n \sim N(\mu, \sigma_2^2)$ where σ_1, σ_2 are unknown, but we have $\sigma_1^2 = \sigma_2^2$. We wish to test

$$\mathcal{H}_0 : \mu_1 - \mu_2 = D_0; \mathcal{H}_1 : \mu_1 - \mu_2 \neq D_0 (\mu_1 - \mu_2 > D_0)(\mu_1 - \mu_2 < D_0)$$

We first compute the value under the null hypothesis:

$$t = \frac{\overline{x}_m - \overline{y}_n - D_0}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}}, s_p^2 = \frac{(m-1)s_m^2 + (n-1)s_n^2}{m+n-2}$$

and we will reject $\mathcal{H}_0$ at significance level α if

$$|t| > t_{\alpha/2,m+n-2}(t > t_{\alpha,m+n-2})(t < -t_{\alpha,m+n-2})$$

Now I have listed a bunch of exercises, do it carefully, and identify which of the above 7 cases! The values of z, t, χ can all be found in the standard table.

Exercise 25. *Atlantic bluefin tuna is the largest and most endangered of the tuna species; the concern is that this species has been overfished and that the mean weight has decreased. Suppose a random sample of 12 Atlantic blue fin tuna was obtained from commercial fishing boats and weighted. The sample is normally distributed with* $x_n = 535.7$ *and* $s_n = 37.8$*. Is there any evidence that the mean weight is less than* 550 *pounds? Use the significance level* $\alpha = 0.05$*.*

A general approach is that, we always make null hypothesis to be simple, i.e it is equal to something. So $\mathcal{H}_0 : \mu = 550$. Also we want to see if its the weight has been reduced, so we make the alternative hypothesis to be $\mathcal{H}_1 : \mu < 550$. Then according to the table, here both μ, σ unknown, we compute the value under the null hypothesis $\mathcal{H}_0$ first:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{12}} = \frac{535.7 - 550}{37.8/\sqrt{12}} = -1.310493468$$

now, in order to reject $\mathcal{H}_0$ under level α, we need $t < -t_{\alpha,n-1}$, which according to the t-table we have $t_{0.05,11} = 1.796$, where we have $-1.31049 > -1.796$, hence we do not reject $\mathcal{H}_0$, so at significance level $\alpha = 0.05$, we do not see any evidence that the mean weight is less than 550 pounds.

Exercise 26. *Despite a sophisticated recycling system, a water park informs the city water department of their need for 1 million liter of water per day. The city water department selected a random sample of* $n = 21$ *days; the mean and sample standard deviation of the park?s water usage (in thousands of liter) were* $x_n = 927.43$*,* $s_n = 154.45$*. Assuming the usage is normally distributed, is there evidence to suggest the mean water usage is different from 1 million liter per day? Use the significance level* $\alpha = 0.05$*.*

Here, we have a similar case as the previous exercise: We're in a normal random sample with both μ, σ^2 unknown. So let $\mathcal{H}_0 : \mu = 1000$ and $\mathcal{H}_1 := \mu \neq 1000$. So again we compute the value under $\mathcal{H}_0$, we have

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{21}} = \frac{927.43 - 1000}{154.45/\sqrt{21}} = -2.153$$

now in order to reject $\mathcal{H}_0$ under level α, we need $|t| > t_{\alpha/2.n-1}$ and from the t-table we have $t_{\alpha/2,n-1} = t_{0.025,20} = 2.086$, and since we do have $|t| = 2.153 > 2.086$, we will hence reject $\mathcal{H}_0$ and in favor of $\mathcal{H}_1$, that is, there is evidence that the mean water usage is different from 1 million liter per day.

Exercise 27. *A study conducted by the Florida Game and Fish Commission aims at assessing the amounts of the DDT insecticide in the brain tissue of brown pelicans. Approximately Normal and independent samples of $n = 10$ juveniles and $m = 13$ nestlings gave (in parts per million), and*

$$\overline{x}_n = 0.041, s_n = 0.017, \overline{y}_m = 0.026, s_m = 0.006.$$

Test whether the mean amounts of DDT in juveniles and nestlings are the same. Use the significance level $\sigma = 0.05$.

In this example, we have two mutually independent random samples, and we design $\mathcal{H}_0 : \mu_1 - \mu_2 = 0$ while $\mathcal{H}_1 : \mu_1 - \mu_2 \neq 0$, and in this two samples we have $\sigma_1^2 = \sigma_2^2$ unknown, so we first compute the value under the null hypothesis, which is

$$t = \frac{\overline{x}_n - \overline{y}_m - 0}{s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}, s_p^2 = \frac{(n-1)s_n^2 + (m-1)s_m^2}{n + m - 2}$$

where we compute $s_p = \sqrt{\dfrac{9 \times 0.017^2 + 12 \times 0.006^2}{10 + 13 - 2}} = 0.012$, and we have

$$t = \frac{0.041 - 0.026}{0.012 \times \sqrt{\frac{1}{10} + \frac{1}{13}}} = 2.97178$$

while we will reject $\mathcal{H}_0$ if $|t| > t_{\alpha/2, n+m-2} = t_{0.025,21} = 2.08$, which we see that clearly we should reject $\mathcal{H}_0$, meaning that the amount of DDT in juveniles and nestlings are not the same.

Exercise 28. *A company produces machine engine parts that are supposed to have a diameter variance no larger than 0.0002. A random sample of $n = 10$ parts gave a sample variance of 0.0003. We wish to test $\mathcal{H}_0 : \sigma^2 = 0.0002$, $\mathcal{H}_1 : \sigma^2 > 0.0002$. at the significance level $\alpha = 0.05$. Assume that the random sample is iid from $N(\mu, \sigma^2)$ with both parameters unknown.*

In this example, we will first compute the value under the null hypothesis $\mathcal{H}_0$:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \times 0.0003}{0.0002} = 13.5$$

In this case we will reject $\mathcal{H}_0$ if we have $\chi^2 > \chi^2_{\alpha, n-1} = \chi^2_{0.05,9} = 16.918$, where we see that we do not satisfy this condition, and hence we will not reject $\mathcal{H}_0$, and we conclude that at significance level $\alpha = 0.05$, we do not have much information that $\sigma^2 > 0.0002$.

Exercise 29. *An experimenter was convinced that the variability in his/her measuring equipment results in a standard deviation of 2; $n = 16$ measurements yielded $s^2 = 6.1$. Do the data disagree with his/her claim? Use the significance level $\alpha = 0.05$. Assume the measurements are normally distributed with both mean and variance unknown.*

In this example, we will test $\mathcal{H}_0 : \sigma^2 = 4$ and $\mathcal{H}_1 : \sigma^2 \neq 4$. *Note that standard deviation is σ!!!.* In this normal sample with both mean and variance unknown, we first compute the value under the null hypothesis $\mathcal{H}_0$:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{15 \times 6.1}{4} = 22.875$$

Under this condition we will reject $\mathcal{H}_0$ if $\chi^2 > \chi^2_{\alpha/2,n-1}$ or $\chi^2 < \chi^2_{1-\alpha/2,n-1}$. Here we have $\alpha = 0.05, n-1 = 15$, and $\chi^2_{0.025,15} = 27.48839$ and $\chi^2_{0.975,15} = 6.26214$, and we see that χ^2 do not satisfy any of these two conditions, so we will not reject $\mathcal{H}_0$, at significance level 0.05.

Exercise 30. *A study published in* 2004 *in Current Allergy and Clinical Immunology concerns the allergy to the powder on latex gloves. Among other things, the exposure to the powder of $n = 46$ hospital employees with diagnosed latex allergy was investigated. The number of latex gloves used per week by these sampled workers is summarized as $\overline{x}_n = 19.3$, $s_n = 11.9$. Is there evidence to conclude that the mean number of latex gloves used per week by hospital employees with latex allergy is more than* 15*? Use $\alpha = 0.01$.*

Here since we have a large sample so we could use a normal approximation for this sample. We will test $\mathcal{H}_0 : \mu = 15$ and $\mathcal{H}_1 : \mu > 15$. We first compute the value under the null hypothesis:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{19.3 - 15}{11.9/\sqrt{46}} = 2.4507$$

We will reject $\mathcal{H}_0$ if $t > t_{\alpha,n-1} = t_{0.01,45} = 2.326$, where we will reject $\mathcal{H}_0$ under this case, at a significance level $\alpha = 0.01$.

Exercise 31. *A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results (in seconds) are summarized as $\overline{x}_m = 3.6, s_m^2 = 0.18$, $\overline{x}_n = 3.8, s_n^2 = 0.14$ Is there evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = 0.05$.*

This one is similiar to the one in exercise 26.

Exercise 32. *A machine in a factory produces* 10% *of defectives among a large lot of items that it produces in a day. A random sample of* 100 *items from the day?s production contains* 15 *defectives, and the supervisor says that the machine must be repaired. Is there evidence that the machine produces more than* 10% *of defectives on average? Use $\alpha = 0.05$.*

Here, we have a Bernoulli random sample, and we test $\mathcal{H}_0 : p = 0.1$ and $\mathcal{H}_1 : p > 0.1$. We first compute the value under the null hypothesis:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.15 - 0.1}{\sqrt{0.1 \times 0.9/100}} = 1.66667$$

and we will reject $\mathcal{H}_0$ if $z > z_\alpha = z_{0.005} = 1.645$, which means we will reject $\mathcal{H}_0$ under significance level $\alpha = 0.05$.

Exercise 33. *Lipitor is a drug that is used to control cholesterol. In a randomized clinical trial,* 94 *subjects were treated with Lipitor and* 270 *independently selected subjects were given a placebo. Among* 94 *treated with Lipitor,* 7 *developed infections, while among* 270 *given a placebo,* 27 *developed infections. Is there a difference between the infection rates for the two drugs? Use* $\alpha = 0.05$.

Here, we have two Bernoulli random samples, let X_i be the sample treated with Lipitor and Y_j be the sample treated with placebo. We have $X_1, \cdots, X_{94} \sim Bernoulli(\theta_1)$, $\theta_1 = 0.074468$ and $Y_1, \cdots, Y_{270} \sim Bernoulli(\theta_2), \theta_2 = 0.1$, we test $\mathcal{H}_0 : \mu_1 - \mu_2 = 0$, and $\mathcal{H}_1 : \mu_1 - \mu_2 \neq 0$. We first compute the value under the null hypothesis:

$$z = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n_1} + \dfrac{\hat{p}(1-\hat{p})}{n_2}}}, \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{7 + 27}{94 + 270} = 0.09341$$

and we get

$$z = \frac{0.074468 - 0.1}{\sqrt{0.09341(1 - 0.09341)/94 + 0.09341(1 - 0.09341)/270}} = -0.73262$$

We will reject $\mathcal{H}_0$ if $|z| > z_{\alpha/2} = z_{0.025} = 1.96$, and hence we do not reject $\mathcal{H}_0$, under the significance level $\alpha = 0.05$.

2 Solved Extra Exercises in Point Estimation

2.1 Does negative matter?

Consider a random sample $X_1, \cdots, X_n \sim f$ where

$$f(x, \theta) = \theta x^{-2}, 0 < \theta \leq x < \infty.$$

(i) Find a sufficient statistics for θ.

(ii) Find the MLE for θ.

(iii) Find the Method of Moment estimator of θ.

Solution :

(i)

The joint pdf of this random sample is given by

$$f(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} \theta x_i^{-2} \chi\{x_i \geq \theta\}$$

where $\chi\{X_i \geq \theta\}$ is the indicator function. Thus it can be simplified to

$$f(\boldsymbol{x}, \theta) = \theta^n \cdot \left(\prod_{i=1}^{n} x_i\right)^{-2} \cdot \chi\{X_{(1)} \geq \theta\}$$

Then by Neyman-Fisher theorem, let

$$g(T(\boldsymbol{X}), \theta) = \theta^n \cdot \chi\{X_{(1)} \geq \theta\}, h(\boldsymbol{x}) = \prod_{i=1}^{n} x_i^{-2}$$

the sufficient statistics is $T(\boldsymbol{X}) = X_{(1)}$.

(ii)

The likelihood function is defined by

$$L_n(\theta) = \theta^n \prod_{i=1}^{n} x_i^{-2} \cdot \chi\{X_{(1)} \geq \theta\}$$

so we can see that if $\theta \leq X_{(1)}$, the function is defined by $L_n(\theta) = K\theta^n$ for some constant K which does not depend on θ, also we can see at this time the likelihood function is increasing; when $\theta \geq X_{(1)}$, because of the indicator function, the likelihood will drop to 0. So the likelihood is maximized at $X_{(1)}$, and thus the MLE for θ is given by $\hat{\theta} = X_{(1)}$.

(iii)

We solve for $\mathbb{E}X_i$ first, which is given by

$$\mathbb{E}X = \int_{\theta}^{\infty} x\theta x^{-2}dx = \theta \ln x \Big|_{x=\theta}^{x=\infty}$$

and we see that $\mathbb{E}X$ is unbounded, thus does not exist, and hence the method of moments estimator does not exist as well.

2.2 Break into Three!

Consider a random sample $X_1, \cdots, X_n$ with parameter $\boldsymbol{\theta} = (\alpha, \beta), \alpha, \beta > 1$ with distribution

$$P_{\boldsymbol{\theta}}(X_i \leq x) = \begin{cases} 0 & x < 0 \\ (x/\beta)^\alpha & 0 \leq x \leq \beta \\ 1 & x > \beta \end{cases}$$

(i) Find a sufficient statistics for $\boldsymbol{\theta}$.

(ii) Find the MLE of $\boldsymbol{\theta}$.

Solution :

(i)

We first differentiate $P_{\boldsymbol{\theta}}$ to get the probability density function. By doing so we have

$$f(x, \theta) = \begin{cases} 0 & x < 0 \\ \alpha x^{\alpha-1}\beta^{-\alpha} & 0 \leq x \leq \beta \\ 0 & x > \beta \end{cases}$$

Then the joint pdf is given by

$$f(\boldsymbol{x}, \theta) = \prod_{i=1}^n \alpha \left(\frac{x_i}{\beta}\right)^{\alpha-1} \cdot \chi\{0 \leq x_i \leq \beta\}$$

where $\chi\{0 \leq x_i \leq \beta\}$ is the indicator function. Furthermore we can write $f(\boldsymbol{x}, \theta)$ as

$$f(\boldsymbol{x}, \theta) = \prod_{i=1}^n \alpha \left(\frac{x_i}{\beta}\right)^{\alpha-1} \cdot \chi\{X_{(1)} \geq 0\} \cdot \chi\{X_{(n)} \leq \beta\}$$

so we let

$$g(T(\boldsymbol{X}), \theta) = \alpha^n \cdot \left(\prod \frac{x_i}{\beta}\right)^{\alpha-1} \cdot \chi\{X_{(n)} \leq \beta\}, h(\boldsymbol{x}) = \chi\{X_{(1)} \geq 0\}$$

then by Neyman-Fisher theorem, we claim that $T(\boldsymbol{X}) = (X_{(n)}, \prod X_i)$ is a sufficient statistics for θ.

(ii)

For β, we note that the likelihood function takes the form

$$L_n(\beta) = K \left(\frac{1}{\beta}\right)^{\alpha-1} \cdot \chi\{X_{(n)} \leq \beta\}$$

where K is a constant that does not depend on β, so we see that when $\beta \leq X_{(n)}$, then $\chi\{X_{(n)} \leq \beta\} = 0$, so the likelihood function is constant 0. When $\beta > X_{(n)}$, we see that since

$\left(\frac{1}{\beta}\right)^{\alpha-1}$ is decreasing, so it will attain its maximum at $X_{(n)}$, and thus the MLE for β is given by $\hat{\beta} = X_{(n)}$.

For α, the likelihood function now takes the form

$$L_n(\alpha) = K \prod_{i=1}^{n} \alpha \left(\frac{x_i}{\beta}\right)^{\alpha-1}$$

for some constant that does not depend on α. Then the log-likelihood is given by

$$\ell_n(\alpha) = n \log \alpha + (\alpha - 1) \log \prod x_i - n(\alpha - 1) \log \beta$$

and we solve for

$$\frac{d\ell_n(\alpha)}{d\alpha} = \frac{n}{\alpha} + \log \prod x_i - n \log \beta = 0$$

we get

$$\hat{\alpha} = \frac{n}{-\log \prod x_i + n \log \beta}$$

where we already computed that $\hat{\beta} = X_{(n)}$, so we get

$$\hat{\alpha} = \frac{n}{-\log \prod_{i=1}^{n} x_i + n \log x_{(n)}}.$$

By second derivative test, we find that

$$\left.\frac{d^2\ell_n(\alpha)}{d\alpha^2}\right|_{\alpha=\hat{\alpha}} < 0$$

so $\hat{\alpha}$ is indeed the MLE of α.

2.3 I am Derivative!

Consider a random sample $X_1, \cdots, X_n$ with pdf

$$f(x, \theta) = \theta x^{\theta - 1}, 0 < x < 1, \theta > 0$$

(i) Find a sufficient statistics for θ.

(ii) Find the CRLB of $\tau(\theta) = \theta$.

(iii) Find the MLE of θ, and show that the variance of MLE goes to 0 as $n \to \infty$.

(iv) Find the method of moment estimator of θ.

Solution:

(i)

The joint pdf can be written as

$$f(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} \theta x^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta - 1}$$

so by Neyman-Fisher theorem, a sufficient statistics is given by

$$T(\boldsymbol{X}) = \prod_{i=1}^{n} X_i.$$

(ii)

Now we have

$$\frac{d}{d\theta} \log f(\boldsymbol{x}, \theta) = \frac{d}{d\theta} \left[n \log \theta + (\theta - 1) \log \prod_{i=1}^{n} x_i \right] = \frac{n}{\theta} + \log \prod_{i=1}^{n} x_i$$

furthermore we take the second derivative, given by

$$\frac{d^2}{d\theta^2} \log f(\boldsymbol{x}, \theta) = -\frac{n}{\theta^2}$$

so the Fisher information is given by

$$\mathcal{I}(\theta) = -\mathbb{E} \left[\frac{d^2}{d\theta^2} \log f(\boldsymbol{x}, \theta) \right] = -\mathbb{E} \left[-\frac{n}{\theta^2} \right] = \frac{n}{\theta^2}.$$

Also we have $[\tau'(\theta)]^2 = 1$, so the CRLB is given by

$$CRLB = \frac{[\tau'(\theta)]^2}{\mathcal{I}(\theta)} = \frac{\theta^2}{n}.$$

(iii)

From (ii), we already know that

$$\frac{d}{d\theta}\ell_n(\theta) = \frac{n}{\theta} + \log \prod_{I=1}^{n} x_i$$

so by setting $\frac{d}{d\theta}\ell_n(\theta) = 0$, we have

$$\hat{\theta}_n = -\frac{n}{\sum_{i=1}^{n} \log x_i},$$

furthermore by second derivative test,we have

$$\left.\frac{d^2}{d\theta^2}\ell_n(\theta)\right|_{\theta=\hat{\theta}} < 0$$

so $\hat{\theta}$ is the MLE. Note that since $X_i \sim f = \theta x^{\theta-1}$, if we let $Y_i = -\log X_i$, then by the theorem of function of random variables, we have

$$\begin{aligned} Y_i = g(X_i) &= f(g^{-1}(Y_i)) \cdot \left| \frac{d}{dy} g^{-1}(Y_i) \right| \\ &= f(e^{-Y}) \cdot \left| \frac{d}{dy} e^{-Y} \right| \\ &= \theta e^{-\theta Y}. \end{aligned}$$

Thus $Y_i \sim Exponential(\theta)$, with pdf $g(y) = \theta e^{-\theta y}$. Then using the property of an exponential distribution, we know that $Y_1 + \cdots + Y_n \sim Gamma(n, \theta)$, so now the MLE is of the form $\hat{\theta} = \frac{n}{Y}$, where $Y \sim Gamma(n, \theta)$. Hence we have

$$\mathbb{E}\left[\frac{n}{Y}\right] = \int_0^\infty \frac{n}{y} \cdot \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y} dy = \frac{n}{n-1}\theta$$

and similarly we can find

$$Var\left(\frac{n}{Y}\right) = \int_0^\infty \frac{n^2}{y^2} \cdot \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y} dy - \left(\mathbb{E}\left[\frac{n}{Y}\right]\right)^2 = \frac{n^2}{(n-1)^2(n-2)^2}\theta^2.$$

Hence the variance of the MLE goes to 0 as $n \to \infty$.

(iv)

We can see that

$$\mathbb{E}X = \int_0^1 x\theta x^{\theta-1} dx = \left.\frac{\theta}{\theta+1} x^{\theta+1}\right|_{x=0}^{x=1} = \frac{\theta}{\theta+1}.$$

Then we set

$$\frac{X_1 + \cdots + X_n}{n} \equiv \frac{\theta}{\theta+1}$$

and we get the estimate to be

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n + \sum_{i=1}^{n} X_i}.$$

2.4 Make it Positive!

Consider a random sample $X_1, \cdots, X_n$ with pdf

$$f(x,\theta) = \frac{1}{2}e^{-|x-\theta|}, x \in \mathbb{R}, \theta \in \mathbb{R}$$

(i) Find the method of moments estimator of θ.

(ii) Find the MLE of θ.

Solution:

(i)

We first find the expected value for X_i, given by

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{\infty} x\frac{1}{2}e^{-|x-\theta|}dx \\
&= \frac{1}{2}\int_{-\infty}^{\theta} xe^{x-\theta}dx + \frac{1}{2}\int_{\theta}^{\infty} xe^{\theta-x}dx \\
&= \frac{1}{2}\left(xe^{x-\theta}\Big|_{-\infty}^{\theta} - \int_{-\infty}^{\theta} e^{x-\theta}dx \right) + \frac{1}{2}\left(-xe^{\theta-x}\Big|_{\theta}^{\infty} - \int_{\theta}^{\infty} -e^{\theta-x}dx \right) \\
&= \frac{1}{2}\left(\theta - e^{x-\theta}\Big|_{-\infty}^{\theta} + \theta - e^{\theta-x}\Big|_{\theta}^{\infty} \right) = \theta.
\end{aligned}$$

So we set

$$\frac{X_1 + \cdots + X_n}{n} \equiv \theta$$

and we get the method of moment estimator to be $\hat{\theta}_n = \overline{X}_n$.

(ii)

The likelihood function is given by

$$L_n(\theta) = \prod_{i=1}^{n} \frac{1}{2}e^{-|x_i-\theta|} = \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} e^{-|x_i-\theta|}.$$

Now, we define the order statistics $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, and then we know there exists integer $1 \leq j \leq n$ such that $x_{(j)} \leq \theta$ and $x_{(j+1)} \geq \theta$, so now we can simplify the likelihood function as

$$
\begin{aligned}
L_n(\theta) &= \left(\frac{1}{2}\right)^n \left(\prod_{x_{(i)} \leq x_{(j)}} e^{-(\theta - x_i)}\right) \cdot \left(\prod_{x_{(i)} \geq x_{(j+1)}} e^{-(x_i - \theta)}\right) \\
&= \left(\frac{1}{2}\right)^n \exp\left\{\sum_{x_{(i)} \leq x_{(j)}} (x_i - \theta) + \sum_{x_{(i)} \geq x_{(j+1)}} (-x_i + \theta)\right\} \\
&= \left(\frac{1}{2}\right)^n \exp\left\{(n-2j)\theta + \sum_{x_{(i)} \leq x_{(j)}} x_i - \sum_{x_{(}i) \geq x_{(j+1)}} x_i\right\}
\end{aligned}
$$

Now we solve

$$
\frac{d}{d\theta} L_n(\theta) = K \exp\left\{(n-2j)\theta + \sum_{x_{(i)} \leq x_{(j)}} x_i - \sum_{x_{(i)} \geq x_{(j+1)}} x_i\right\}(n-2j) = 0
$$

where K is a positive constant that does not depend on θ, and since the exponential is always positive, we have $n = 2j$. When n is even, we have $j = \frac{n}{2}$, meaning $X_{(\frac{n}{2})}$ is the MLE, when n is odd, then we take the "mid-point", given by $X_{(\frac{n+1}{2})}$.

2.5 I Look Scary!

Suppose X has a pdf

$$f(x,\theta) = \left(\frac{\theta}{2}\right)^{|x|} \cdot (1-\theta)^{1-|x|}, x = -1, 0, 1, 0 \leq \theta \leq 1,$$

Show that the estimator

$$T(X) = \begin{cases} 2 & x = 1 \\ 0 & \text{otherwise} \end{cases},$$

is an unbiased estimator of θ, and find a better estimator of θ.

Solution: In $T(X)$, we have

$$\mathbb{E}[T(X)] = 2 \times P(X = 1) + 0 \times P(X = 0, -1) = 2 \times \left(\frac{\theta}{2}\right) = \theta.$$

Hence $T(X)$ is unbiased. Next we consider the MLE of θ, when $x = 1, -1$, we have $L(\theta) = \frac{\theta}{2}$, and we see that the likelihood function is increasing on $[0, 1]$, and hence the maximum likelihood is when $\theta = 1$, given by $\hat{\theta} = 1$. When $x = 0$, we have the likelihood function is $L(\theta) = 1 - \theta$, and we see when $\theta = 0$, the likelihood function is maximized and hence $\hat{\theta} = 0$, so the MLE is given by

$$MLE(\theta) = \begin{cases} 1 & x = 1 \\ 1 & x = -1 \\ 0 & x = 0 \end{cases}$$

We can easily varify that the MLE is unbiased, since

$$\mathbb{E}[MLE(\theta)] = 1 \times P(X = 1) + 1 \times P(X = -1) + 0 \times P(X = 0) = \frac{\theta}{2} + \frac{\theta}{2} = \theta.$$

The variance of the MLE is

$$\begin{aligned} Var[MLE(\theta)] &= \mathbb{E}[MLE^2(\theta)] - \theta^2 \\ &= 1^2 \times P(X = 1) + 1^2 \times P(X = -1) + 0^2 \times P(X = 0) - \theta^2 \\ &= \theta - \theta^2. \end{aligned}$$

Also

$$\begin{aligned} Var(T(X)) &= \mathbb{E}[T^2(X)] - \theta^2 \\ &= 4 \times P(X = 1) + 0 \times P(X = 0) - \theta^2 \\ &= 2\theta - \theta^2. \end{aligned}$$

When $0 \leq \theta \leq 1$, we see that $Var(MLE(\theta)) \leq Var(T(X))$, so the MLE is indeed a better estimator of θ.

2.6 Am I Normal?

Consider a random sample $Y_1, \cdots, Y_n$ where

$$Y_i = \beta x_i + \epsilon_i$$

where $\epsilon_i \sim N(0, \sigma^2)$. The parameters are β and σ^2, and $x_1, \cdots, x_n$ are fixed constants.

(i) Find a sufficient statistics for (β, σ^2).

(ii) Find the MLE of β, and show the MLE is unbiased.

(iii) Furthermore, let

$$W_n = \frac{\sum Y_i}{\sum x_i}, Z_n = \frac{\sum Y_i/x_i}{n}$$

then show that both W_n, Z_n are unbiased, and compute both variances.

Solution:

(i)

First note that $Y_i \sim N(\beta x_i, \sigma^2)$, then the joint pdf of Y_i is given by

$$\begin{aligned} f(\boldsymbol{Y}, \boldsymbol{\theta}) &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot e^{-\frac{(y_i-\beta x_i)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta x_i)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{1}{\sigma^2}\sum \beta x_i y_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\beta^2 x_i^2\right\} \end{aligned}$$

where if we let $T(\boldsymbol{Y}) = \left(\sum_{i=1}^{n} x_i Y_i, \sum_{i=1}^{n} Y_i^2\right)$, then by Neyman-Fisher theorem, we know that $T(\boldsymbol{Y})$ is a sufficient statistics for (β, σ^2).

2.7 Multiplier!

Consider a random sample $X_1, \cdots, X_n \sim Bernoulli(p)$, $0 < p < 1$.

(i) Show that the MLE of p attains the Cramér-Rao lower bound.

(ii) For $n \geq 4$, find the UMVUE of $\tau(p) = p^4$.

Solution :

(i)

First of all we find the MLE of p: The likelihood function is defined as

$$L_n(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

so the log-likelihood is

$$\ell_n(p) = n\overline{x}_n \log p + (n - n\overline{x}_n)\log(1-p)$$

we take the derivative w.r.t p:

$$\frac{d\ell_n(p)}{dp} = \frac{n\overline{x}_n}{p} - \frac{n - n\overline{x}_n}{1-p}$$

by solving $\dfrac{d\ell_n(p)}{dp} = 0$, we get $\hat{p} = \overline{x}_n$, since furthermore

$$\left.\frac{d^2\ell_n(p)}{dp^2}\right|_{p=\hat{p}} < 0$$

so $\hat{p} = \overline{x}_n$ is indeed the MLE. Note that

$$\frac{d\ell_n(p)}{dp} = \frac{n}{p(1-p)}(\overline{x}_n - p)$$

where we let $a(p) = \dfrac{n}{p(1-p)}$, $T(\boldsymbol{X}) = \overline{X}_n$, with $\mathbb{E}[T(\boldsymbol{X})] = p$ unbiased, so by the theorem about CRLB, we claim that $T(\boldsymbol{X}) = \overline{x}_n$ attains the CRLB.

(ii)

By independence, we can easily see that $X_1X_2X_3X_4$ is an unbiased estimator for p^4. Now since the joint pdf of the random sample is given by

$$f(\boldsymbol{x}, p) = p^{n\overline{x}_n}(1-p)^{n-n\overline{x}_n} = \left(\frac{p}{1-p}\right)^{n\overline{x}_n} \cdot (1-p)^n$$

so by Neyman-Fisher theorem, $T(\boldsymbol{X})$ is a sufficient statistics, and furthermore it is complete. So by Rao-Blackwell theorem, the UMVUE for p^4 is given by

$$\begin{aligned}
\delta(\boldsymbol{X}) &= \mathbb{E}\left[X_1X_2X_3X_4 \middle| \sum_{i=1}^n X_i = t\right] \\
&= 1 \times P\left(X_1X_2X_3X_4 = 1 \middle| \sum_{i=1}^n X_i = t\right) + 0 \times P\left(X_1X_2X_3X_4 = 0 \middle| \sum_{i=1}^n X_i = t\right) \\
&= \frac{P\left(X_1X_2X_3X_4 = 1, \displaystyle\sum_{i=1}^n X_i = t\right)}{P\left(\displaystyle\sum_{i=1}^n X_i = t\right)} \\
&= \frac{p^4 \cdot \dbinom{n-4}{t-4} p^{t-4}(1-p)^{n-t}}{\dbinom{n}{t} p^t (1-p)^{n-t}} \\
&= \frac{\dbinom{n-4}{t-4}}{\dbinom{n}{t}}
\end{aligned}$$

where $t = \displaystyle\sum_{i=1}^n X_i$. This is the UMVUE of $\tau(\theta) = p^4$.

Remark: *It is completely doable if I make* $\tau(\theta) = p^{2025}$*!*

2.8 Sum or no?

Consider a random sample $X_1, \cdots, X_n \sim Bernoulli(p)$, then define

$$h(p) = P\left(\sum_{i=1}^{n-1} X_i > X_n\right),$$

find the UMVUE of $h(p)$.

Solution :

First we define the estimator as the indicator function

$$T(\boldsymbol{X}) = \chi\left\{\sum_{I=1}^{n} X_i > X_{n+1}\right\} = \begin{cases} 1 & X_1 + \cdots + X_n > X_{n+1} \\ 0 & \text{otherwise} \end{cases}$$

It is easy to show $T(\boldsymbol{X})$ is unbiased, since

$$\mathbb{E}[T(\boldsymbol{X}) = 1 \times P(T(\boldsymbol{X}) = 1) + 0 \times T(\boldsymbol{X}) = 0) = h(p).$$

Then from **Q7** we already showed that $\sum_{i=1}^{n} X_i$ is a complete sufficient statistics of p, then by Rao-Blackwell theorem, the UMVUE s given by

$$\begin{aligned}
\delta(\boldsymbol{X}) &= \mathbb{E}\left[h(p)\middle|\sum_{i=1}^{n} X_i = t\right] \\
&= P\left(\sum_{i=1}^{n} X_i > X_{n+1}\middle|\sum_{i=1}^{n} X_i = t\right) \\
&= \frac{P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n} X_i = t\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n} X_i = t\right)}{\binom{n}{t} p^t (1-p)^{n-t}}
\end{aligned}$$

We will now investigate the numerator, denote the numerator to be N

If $t = 0$, then $N = 0$;

If $t = 1$, then

$$N = \binom{n}{1} p(1-p)^{n-1} \times (1-p)$$

If $t \geq 2$, then

$$N = \binom{n}{t} p^t (1-p)^{n-t} \times (1-p) + \binom{n}{t} p^t (1-p)^{n-t} \times p$$

So in all, we have

$$\delta(\boldsymbol{X}) = \begin{cases} 0 & t = 0 \\ \\ \dfrac{\binom{n}{1} p (1-p)^{n-1} \times (1-p)}{\binom{n}{t} p^t (1-p)^{n-t}} = \dfrac{\binom{n}{1}}{\binom{n}{t}} & t = 1 \\ \\ \dfrac{\binom{n}{t} p^t (1-p)^{n-t} \times (1-p) + \binom{n}{t} p^t (1-p)^{n-t} \times p}{\binom{n}{t} p^t (1-p)^{n-t}} = 1 & t \geq 2 \end{cases}$$

where $t = \sum_{i=1}^{n} X_i$, and this is the UMVUE.

2.9 I am not weak!

Consider the random sample $X_1, \cdots, X_n \sim Uniform(0, \theta)$, and the prior distribution for θ is given by

$$\pi(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}, \alpha, \beta > 0; \theta > 1$$

Find the Bayes estimator of θ under the squared error loss.

Solution: Recall that the posterior distribution takes the form

$$\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\displaystyle\int_\Theta p(x|\theta)\pi(\theta)d\theta}$$

where

$$\begin{aligned} p(x|\theta)\pi(\theta) &= \left(\frac{1}{\theta}\right)^n \cdot \chi\{X_{(1)} > 0\}\chi\{X_{(n)} < \theta\} \cdot \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} \cdot \chi\{\theta > 1\} \\ &= \theta^{-(n+\alpha+1)} \cdot \alpha\beta^\alpha \cdot \chi\{\theta > \max\{1, X_{(n)}\}\} \cdot \chi\{X_{(1)} > 0\}. \end{aligned}$$

So it follows that

$$\begin{aligned} \int_\Theta p(x|\theta)\pi(\theta)d\theta &= \int_{T(\boldsymbol{X})}^\infty \theta^{-(n+\alpha+1)}\alpha\beta^\alpha\chi\{X_{(1)} > 0\}d\theta \text{ where } T(\boldsymbol{X}) = \max\{1, X_{(n)}\} \\ &= \chi\{X_{(1)} > 0\}\alpha\beta^\alpha \left[-\frac{1}{n+\alpha}\theta^{-(n+\alpha)}\right]\Bigg|_{\theta=T(\boldsymbol{X})}^\infty \\ &= \chi\{X_{(1)} > 0\}\alpha\beta^\alpha \left[\frac{1}{n+\alpha}(T(\boldsymbol{X}))^{-(n+\alpha)}\right] \end{aligned}$$

so now the posterior distribution is the ratio of what we get, and we can simplify to

$$\pi(\theta|x) = (n+\alpha) \cdot \left(\frac{1}{\theta}\right)^{n+\alpha+1} \cdot [T(\boldsymbol{X})]^{-(n+\alpha)}; \theta > T(\boldsymbol{X}).$$

Now, under the squared error loss, the Bayes estimator of θ is given by the expected value of the posterior distribution, hence

$$\begin{aligned} \hat{\theta}_{\text{Bayes}} = \mathbb{E}[\pi(\theta|x)] &= \int_{T(\boldsymbol{X})}^\infty \theta \cdot (n+\alpha) \cdot \theta^{-(n+\alpha+1)} \cdot [T(\boldsymbol{X})]^{-(n+\alpha)}d\theta \\ &= (n+\alpha) \cdot [T(\boldsymbol{X})]^{-(n+\alpha)} \int_{T(\boldsymbol{X})}^\infty \theta^{-(n+\alpha)}d\theta \\ &= \frac{n+\alpha}{-(n+\alpha)+1}[T(\boldsymbol{X})]^{-(n+\alpha)}[\theta^{-(n+\alpha)+1}]\Bigg|_{\theta=T(\boldsymbol{X})}^\infty \\ &= \frac{n+\alpha}{n+\alpha-1}T(\boldsymbol{X}). \end{aligned}$$

2.10 Interchanging Normal!

Consider the random sample $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$. In each of the following cases, find the CRLB as well as the UMVUE for the parameter of interest. Compute the variance of the UMVUE.

(i) Suppose now we have $\sigma^2 = 1$, our parameter of interest is $\tau(\mu) = \mu^2$.

(ii) Suppose now we have $\mu = 0$, our parameter of interest is $\tau(\sigma) = \sigma, \sigma > 1$.

Solution:

(i)

First, we find the Fisher imformation $\mathcal{I}_1(\mu)$. Note that the log-based pdf is

$$\log f(x, \mu) = \log \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = \log \frac{1}{\sqrt{2\pi}} - \frac{(x-\mu)^2}{2}$$

Hence we have

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

so using the property of Fisher information as well as Bartlett's identity, we claim that

$$\mathcal{I}_n(\mu) = n\mathcal{I}_1(\mu) = -n\mathbb{E}\left\{\frac{\partial^2 \log f(x, \mu)}{\partial \mu^2}\right\} = n.$$

Now, we have $[\tau'(\mu)]^2 = 4\mu^2$, so the CRLB is given by

$$CRLB = \frac{4\mu^2}{n}.$$

We know that the sufficient statistic of μ is given by the sample mean, $\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$, and thus under transformations, the sufficient statistic of μ^2 is just $T(\boldsymbol{X}) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \overline{X}_n^2$. Also we know that $T(\boldsymbol{X})$ is complete under the exponential family, so we would like to apply Lehmann-Scheffé theorem and aim to find an unbiased estimator which is a function of $T(\boldsymbol{X})$. Note that it is easy to verify that $\overline{X}_n \sim N\left(\mu, \frac{1}{n}\right)$, and hence

$$\mathbb{E}T(\boldsymbol{X}) = \mathbb{E}\overline{X}_n^2 = Var(\overline{X}_n) + (\mathbb{E}\overline{X}_n)^2 = \frac{1}{n} + \mu^2.$$

So an obvious choice of the unbiased estimator of θ would be $U(\boldsymbol{X}) = \overline{X}_n^2 - 1/n$. Then by Lehmann-Scheffé theorem, $U(\boldsymbol{X})$ is the UMVUE of $\tau(\mu) = \mu^2$, since $\overline{X}_n^2$ is a complete sufficient statistic of μ^2. To compute $Var(\overline{X}_n^2)$, we will use the relation between normal and chi-squared distribution. Observe that

$$
\begin{aligned}
Var\left(\overline{X}_n^2\right) &= Var\left(\frac{1}{n}\left(\frac{\overline{X}_n-\mu}{\sqrt{\frac{1}{n}}}\right)^2 - 2\mu\overline{X}_n - \mu^2\right) \\
&= \frac{1}{n^2}Var(\chi_1^2) + 4\mu^2 Var(\overline{X}_n) \\
&= \frac{2}{n^2} + \frac{4\mu^2}{n}.
\end{aligned}
$$

(ii)

Here, the log-pdf takes the form

$$
\log f(x,\sigma) = \log\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{x^2}{2\sigma}} = \log\frac{1}{\sqrt{2\pi}} - \log\sigma - \frac{x^2}{2\sigma}.
$$

So the second partial derivative is given by

$$
\frac{\partial^2\log f(x,\sigma)}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{x^2}{\sigma^3}
$$

Hence the Fisher information is given by

$$
\mathcal{I}_1(\sigma) = -\mathbb{E}\left[\frac{1}{\sigma^2} - \frac{X^2}{\sigma^3}\right] = -\frac{1}{\sigma^2} + \frac{1}{\sigma^3}(Var(X) + [\mathbb{E}X]^2) = \frac{1}{\sigma} - \frac{1}{\sigma^2}, \sigma > 1.
$$

Given that $[\tau(\sigma)]' = 1$, we have the CRLB given by

$$
CRLB = \frac{1}{n\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2}\right)}.
$$

Also in this family the sufficient statistic of σ^2 is given by S_n^2, then the sufficient statistic of σ is just $T(\boldsymbol{X}) = \sqrt{S_n^2}$. Since it is also complete under the exponential family, so we would like to use Lehmann-Scheffé and design an unbiased estimator based on $T(\boldsymbol{X})$. We use the fact that

$$
\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2, \sqrt{\chi_n^2} \sim \chi_n
$$

Hence we have

$$
\mathbb{E}\left[\sqrt{\frac{(n-1)S_n^2}{\sigma^2}}\right] = \mathbb{E}\chi_{n-1} = \sqrt{2}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}
$$

Which means we have

$$
\mathbb{E}\left[\sqrt{S_n^2}\right] = \frac{\sqrt{2}}{\sqrt{n-1}} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \cdot \sigma,
$$

so an obvious unbiased estimator would be

$$
U(\boldsymbol{X}) = \frac{\sqrt{n-1}}{\sqrt{2}} \cdot \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \cdot \sqrt{S_n^2}
$$

Then by Lehmann-Scheffé, $U(\boldsymbol{X})$ is indeed the UMVUE of σ. Now to find the variance, note that we have

$$\begin{aligned} Var\left(\sqrt{S_n^2}\right) &= \mathbb{E}[S_n^2] - \left[\mathbb{E}\left(\sqrt{S_n^2}\right)\right]^2 \\ &= \sigma^2 - \left[\frac{\sqrt{2}}{\sqrt{n-1}} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \cdot \sigma\right]^2 \end{aligned}$$

Hence we have

$$Var(U(\boldsymbol{X})) = \left[\frac{\sqrt{n-1}}{\sqrt{2}} \cdot \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}\right]^2 \cdot \left\{\sigma^2 - \left[\frac{\sqrt{2}}{\sqrt{n-1}} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \cdot \sigma\right]^2\right\}.$$

2.11 Transforming an Exponential!

Consider a random sample $X_1, \cdots, X_n \sim Exponential(\theta)$ with $\mathbb{E}X_i = \theta$. That is, the pdf is $f_X(x) = \frac{1}{\theta}e^{-x/\theta}$.

(i) Find the distribution of the MLE of θ.

(ii) Assume $n = 8$, denote $\hat{\theta}$ as the MLE of θ, find $P(\hat{\theta} > 2\theta)$.

(iii) Let $\hat{F}(x, \theta)$ be the MLE of $P(X_1 \leq x)$ for given x, find an approximate expression for $Var[\hat{F}(x, \theta)]$.

2.12 Uniform but not Ordinary!

Suppose we have a random sample $X_1, \cdots, X_n \sim Uniform(0, \theta)$ where $\theta > 0$.

(i) If we are interested in testing $\tau(\theta) = \theta$, what is a sufficient statistic?

(ii) Suggest an unbiased estimator of θ, and hence find the UMVUE of θ. Furthermore, find the UMVUE of θ^r for any positive integer $r < n$.

(iii) Find the variance of the UMVUE of θ. It turns out that the CRLB in this case is $\dfrac{\theta^2}{n}$. What do you see, and why?

Solution:

(i)

It is not hard to derive the joint pdf as $f(\boldsymbol{x}, \theta) = \left(\frac{1}{\theta}\right)^n \cdot \chi\{X_{(1)} > 0\} \cdot \chi\{X_{(n)} < \theta\}$. So by the factorization theorem, a sufficient statistic is given by $T(\boldsymbol{X}) = X_{(n)}$.

(ii)

We suggest the unbiased estimator based on the sufficient statistics: We can find the cdf of $X_{(n)}$:

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = \Big[P(X_i \leq x)\Big]^n = \left[\int_0^x \frac{1}{\theta} dx\right]^n = \frac{x^n}{\theta^n}.$$

Differentiating will yield $f_{X_{(n)}}(x)$:

$$f_{X_{(n)}}(x) = \frac{d}{dx}\frac{x^n}{\theta^n} = \frac{nx^{n-1}}{\theta^n}.$$

Now we have

$$\mathbb{E}[X_{(n)}] = \int_0^\theta x \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{\theta^n} \cdot \frac{1}{n+1} x^{n+1}\Bigg|_0^\theta = \frac{n}{n+1}\theta.$$

Hence an obvious choice of unbiased estimator would be $U(\boldsymbol{X}) = \frac{n+1}{n} X_{(n)}$. Also since $X_{(n)}$ is a complete sufficient statistic, by Lehmann-Scheffé we know that $U(\boldsymbol{X})$ is indeed the UMVUE.

Now, we investigate the second moment:

$$\mathbb{E}[X_{(n)}^2] = \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1} dx = \frac{n}{\theta^n} \cdot \frac{1}{n+2} x^{n+2}\Bigg|_0^\theta = \frac{n}{n+2}\theta^2.$$

Hence an obvious choice of unbiased estimator would be $U_2(\boldsymbol{X}) = \frac{n+2}{n} X_{(n)}^2$, and again by Lehmann-Scheffé theorem, it is the UMVUE. We could generalize this idea, and we investigate the rth moment:

$$\mathbb{E}[X_{(n)}^r] = \int_0^\theta x^r \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{\theta}\int_0^\theta x^{n+r-1} dx = \frac{n}{n+r}\theta^r.$$

So again, we conclude the UMVUE of θ^r is $\frac{n+r}{n} X_{(n)}^r$.

(iii)

The variance of $U(\boldsymbol{X})$ is given by

$$Var(U(\boldsymbol{X})) = \mathbb{E}U^2(\boldsymbol{X}) - [\mathbb{E}U(\boldsymbol{X})]^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2.$$

The variance is smaller than CRLB, this is because the uniform family does not satisfy the regularity conditions! (The support does depend on θ).

2.13 Adding them makes perfect!

Consider a random sample of geometric distribution, that is, $X_1, \cdots, X_n \sim p(1-p)^{x-1}$, $0 < p < 1$, $x = 1, 2, \cdots$.

(i) Find the method of moment estimator of p.

(ii) What is $\mathbb{E}\left[\frac{1}{\overline{X}_n}\right]$?

(iii) What is the UMVUE of p?

Solution:

(i)

In fact we know that $\mathbb{E}X = \frac{1}{p}$ *from the table, but what if Khalili asks you to prove it? I will do the derivations.*

By definition, we have

$$\mathbb{E}X = \sum_{X=1}^{\infty} Xp(1-p)^{X-1} = p\Big(1 + 2(1-p) + 3(1-p)^2 + \cdots\Big)$$

$$(1-p)\mathbb{E}X = p\Big((1-p) + 2(1-p)^2 + 3(1-p)^3 + \cdots\Big)$$

Hence

$$p\mathbb{E}X = p\Big(1 + (1-p) + (1-p)^2 + \cdots\Big) = p\frac{1}{1-(1-p)} = 1.$$

which means $\mathbb{E}X = \frac{1}{p}$, and so we match $\overline{X}_n = \dfrac{1}{p}$, and we see that $\hat{p} = 1/\overline{X}_n$.

(ii) **Skip this, Khalili is so wrong on this**

It is not hard to see that I am wrong, the sum should be negative binomial! By definition we have

$$\mathbb{E}\left[\frac{1}{\overline{X}_n}\right] = \sum_{x=1}^{\infty} \frac{1}{x} \cdot p(1-p)^{x-1} = p\sum_{x=0}^{\infty} \frac{(1-p)^x}{x+1}.$$

In fact, it is a well known Taylor series, and it is easy to see that

$$p\sum_{x=0}^{\infty} \frac{(1-p)^x}{x+1} = \frac{p\ln p}{p-1}.$$

What are you doing Khalili? You are the one who designed this problem.

(iii)

First, define the estimator $T(\boldsymbol{X}) = \begin{cases} 1 & X = 1 \\ 0 & \text{otherwise} \end{cases}$, and it is easy to see that $T(\boldsymbol{X})$ is unbiased, since

$$\mathbb{E}[T(\boldsymbol{X})] = 1 \times P(X = 1) + 0 = p(1-p)^0 = p.$$

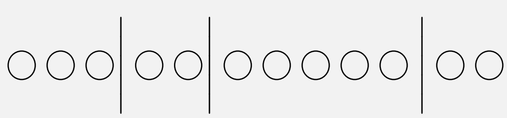
Also it is easy to see that a complete sufficient statistic takes the form $\sum_{i=1}^n X_i$, so we apply Rao-Blackwell theorem, the UMVUE $U(\boldsymbol{X})$ of p takes the form

$$\begin{aligned} U(\boldsymbol{X}) = \mathbb{E}\Big\{T(X)\Big|\sum_{i=1}^n X_i = t\Big\} &= P\Big\{X_1 = 1\Big|\sum_{i=1}^n X_i = t\Big\} \\ &= \frac{P\left\{X_1 = 1, \sum_{i=1}^n X_i = t\right\}}{P\left\{\sum_{i=1}^n X_i = t\right\}} \\ &= \frac{\binom{t-2}{n-2}}{\binom{t-1}{n-1}} = \frac{n-1}{t-1}, \end{aligned}$$

where $t = \sum_{i=1}^n X_i$ is the sufficient statistic. We can do that, because $X_1 | \sum X_i \sim Uniform(1, 2, \cdots, n)$.

Remark: (A part specifically designed for Suelynn and Henry since they do not find the deviation to be trivial).

How do we get the quantity $\binom{t-1}{n-1}$ and the other one? Well if we look at the denominator first, it means, if we randomly assign numbers to each X_i, what is the probability that their sum is t? The same idea applies for the numerator, so we see that we can cancel out the denominator of the probability, hence we only work with the number of cases such that $\sum_{i=1}^n X_i = t$. Also each X_i is at least one, so we look the following diagram, suppose we have t balls, and we use bars to separate them into n pieces, like the description $\sum_{i=1}^n X_i = t$. (The diagram below shows a case when $t = 12, n = 4$.



Since each separation must be at least 1, so our bar can only be placed on the 11 gaps in the middle. Also in order to create 4 parts, we only need 3 bars. So in general it simplified to the case where we place $n-1$ bars on $t-1$ positions, that's why we have $\binom{t-1}{n-1}$. Then you can do the same for the numerator, where specify one piece must be 1 ($\sum_{i=1}^n X_i = t, X_1 = 1$), so it means we have one fixed choice, and we reduce it to $t-1$ balls and $n-1$ bars, and the same idea applies.

Also here, we are saying $X_i \geq 1$ for all i, what if I remove this restriction? What if exactly k of those pieces must be non-zero? In fact, this is your Q1 on MATH 356 final exam.

2.14 I am mean, but in a different way...

Consider a random sample $X_1, \cdots, X_n \sim f(x, \theta)$ where

$$f(x, \theta) = \left(\frac{1}{\theta}\right) r y^{r-1} e^{-y^r/\theta}, \theta > 0, y > 0$$

for some **known** positive constant r.

(i) Derive a sufficient statistic of θ;

(ii) Is the MLE of θ indeed the UMVUE?

3 Solved Extra Exercises in Hypothesis Testing

3.1 Pareto!

A random sample $X_1, \cdots, X_n$ has the following distribution:

$$f(x) = \begin{cases} \dfrac{\theta\nu^\theta}{x^{\theta+1}} & x \geq \nu \\ \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0, \nu > 0$ are parameters.

(i) Find the MLE for θ and ν;

(ii) Given the test $\mathcal{H}_0 : \theta = 1, \mathcal{H}_1 : \theta \neq 1$ for unknown ν, find its critical region.

Solution:

(i)

We first derive the joint pdf, given by

$$f(\boldsymbol{x}) = (\theta\nu^\theta)^n \prod_{i=1}^{n} \frac{1}{x^{\theta+1}} \cdot \chi\{x_i \geq \nu\} = (\theta\nu^\theta)^n \prod_{i=1}^{n} \frac{1}{x^{\theta+1}} \cdot \chi\{x_{(1)} \geq \nu\}$$

Now we investigate the likelihood function. For ν, we see that $L_n(\nu) = 0$ when $\nu \geq X_{(1)}$, and when $0 < \nu < X_{(1)}$, the function $L_n(\nu) = K\nu^{\theta n}$ is increasing. So the MLE for ν is given by $\hat{\nu}_{MLE} = x_{(1)}$. For θ, we investigate the log-likelihood:

$$\ell_n(\theta) = n \log\theta + n\theta \log\nu - \theta \sum_{i=1}^{n} \log x_i + K$$

where K is a constant that does not depend on θ, and hence we solve

$$\frac{\partial \ell_n(\theta)}{\partial\theta} = \frac{n}{\theta} + n \log\nu - \log \prod_{i=1}^{n} x_i = 0$$

we get $\hat{\theta} = \dfrac{n}{\log \prod x_i - n \log x_{(1)}}$, where we replaced ν by its MLE. Furthermore, we have

$$\left.\frac{\partial^2 \ell_n(\theta)}{\partial\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n}{\hat{\theta}^2} < 0$$

so $\hat{\theta}$ is indeed the maximum, and hence the MLE for θ is $\hat{\theta}_{MLE} = \dfrac{n}{T}$ where

$$T = \log\left[\frac{\prod x_i}{x_{(1)}^n}\right].$$

(ii)

Given $\mathcal{H}_0 : \theta = 1$, so the maximum likelihood over $\mathcal{H}_0$ is $\hat{\theta}_{MLE,\mathcal{H}_0} = 1$ and $\hat{\nu}_{MLE,\mathcal{H}'} = x_{(1)}$. Similarly for the entire sample space Θ, the MLEs are the ones we obtained in (i). So by LRT, it's easy to see that

$$\hat{\lambda}_n = \frac{L_n(\hat{\boldsymbol{\theta}}_{MLE,\mathcal{H}_0})}{L_n(\hat{\boldsymbol{\theta}}_{MLE,\Theta})} = \frac{x_{(1)}^n \prod \frac{1}{x_i^2}}{(\frac{n}{T} x_{(1)}^{n/T})^n \prod \frac{1}{x^{n/T+1}}} = \left(\frac{T}{n}\right)^n \frac{e^{-T}}{(e^{-T})^{n/T}} = \left(\frac{T}{n}\right)^n e^{-T+n}.$$

when $\hat{\lambda}_n < C$ for some constant $0 \le C \le 1$, we would reject $\mathcal{H}_0$ and in favour of $\mathcal{H}_1$, and hence the critical region is given by

$$\mathcal{R} := \left\{ \left(\frac{T}{n}\right)^n e^{-T+n} < C, C \in [0,1] \right\}$$

which is equivalent to say that $\mathcal{R} := \{T < c_1\} \bigcup \{T > c_2\}$ for some $c_1 < c_2$, according to the graph of the function $f(x) = (x/n)^n e^{-x+n}$.

3.2 Two Exponentials!

Consider two mutually independent random samples, where $X_1, \cdots, X_n \sim \textit{Exponential}(\theta)$ and $Y_1, \cdots, Y_m \sim \textit{Exponential}(\mu)$. Consider the test $\mathcal{H}_0 : \theta = \mu$ and $\mathcal{H}_1 : \theta \neq \mu$, by using LRT, find its critical region.

Solution:

We first derive the joint pdf, given by

$$f(\boldsymbol{x}, \boldsymbol{y}) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \cdot \mu^m e^{-\mu \sum_{j=1}^m y_j} = \theta^n \mu^m \cdot \exp\left\{-\theta \sum_{i=1}^n x_i - \mu \sum_{j=1}^m y_j\right\}$$

By the property of exponential distribution, it is easy to show that the MLE for θ and μ are

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^n X_i} \qquad \hat{\mu}_{MLE} = \frac{m}{\sum_{j=1}^m Y_j}$$

when $\theta = \mu$, the likelihood function now becomes

$$L_n(\theta) = \theta^{n+m} \cdot \exp\left\{-\theta\left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j\right)\right\}$$

and hence

$$\ell_n(\theta) = (n+m)\log\theta - \theta\left(\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j\right) \quad \text{and} \quad \frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{n+m}{\theta} - \left(\sum_{i=1}^n X_i + \sum_{j=1}^n Y_j\right)$$

where we can also easily see that the MLE is given by

$$\hat{\theta}_{MLE,\theta=\mu} = \frac{n+m}{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}$$

Now, using LRT, the likelihood ratio is given by

$$\hat{\lambda}_n = \frac{L_n(\hat{\theta}_{MLE,\mathcal{H}_0}, \hat{\mu}_{MLE,\mathcal{H}_0})}{L_n(\hat{\theta}_{MLE,\Theta}, \hat{\mu}_{MLE,\Theta})} = \frac{\left(\frac{m+n}{\sum X_i + \sum Y_j}\right)^{n+m}}{\left(\frac{n}{\sum X_i}\right)^n \left(\frac{m}{\sum Y_j}\right)^m}$$

Let $T = \dfrac{\sum X_i}{\sum X_i + \sum Y_j}$, then it is easy to see that

$$\hat{\lambda}_n = \frac{(m+n)^{m+n}}{n^n \cdot m^m} T^n (1-T)^m.$$

Hence the critical region takes the form

$$\mathcal{R} = \{T < a\} \bigcup \{T > b\}.$$

Also, we know that the rejection region takes the form (under regularity conditions)

$$\mathcal{R} := \left\{-2\log\hat{\lambda}_n \geq \chi^2_{\alpha,d}\right\}$$

where α is the significance level, and here $d = 2$ since the parameter is now 2 dimensional.

3.3 Two Betas!

Now, consider two mutually independent random samples, where $X_1, \cdots, X_n \sim Beta(\mu, 1)$ and $Y_1, \cdots, Y_m \sim Beta(\theta, 1)$, where $\mu, \theta \in \mathbb{N}$. Again we are interested in the test $\mathcal{H}_0 : \mu = \theta$ and $\mathcal{H}_1 : \mu \neq \theta$, and find its critical region.

Solution:

Firstly, we derive the joint pdf, given by

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}) &= \left(\frac{\Gamma(\mu+1)}{\Gamma(\mu)}\right)^n \left(\prod_{i=1}^n x_i\right)^{\mu-1} \cdot \left(\frac{\Gamma(\theta+1)}{\Gamma(\theta)}\right)^m \left(\prod_{j=1}^m y_j\right)^{\theta-1} \\ &= \mu^n \left(\prod_{i=1}^n x_i\right)^{\mu-1} \cdot \theta^m \left(\prod_{j=1}^m y_j\right)^{\theta-1} \end{aligned}$$

where the log-likelihood function is given by

$$\ell_n(\mu, \theta) = n \log \mu + (\mu - 1) \log \left(\prod_{i=1}^n x_i\right) + m \log \theta + (\theta - 1) \log \left(\prod_{j=1}^m y_j\right)$$

Hence, it is easy to see that the MLE for μ, θ are given by

$$\hat{\mu}_{MLE} = -\frac{n}{\log \prod x_i} \qquad \hat{\theta}_{MLE} = -\frac{m}{\log \prod y_j}$$

Now under the condition $\mathcal{H}_0 : \theta = \mu$, the log-likelihood is now

$$\ell_n(\theta) = (n+m) \log \theta + (\mu - 1) \left(\log \prod x_i + \log \prod y_j\right)$$

Hence it is also easy to see that the MLE is

$$\hat{\mu}_{MLE,\mathcal{H}_0} = \hat{\theta}_{MLE,\mathcal{H}_0} = -\frac{n+m}{\log \prod x_i + \log \prod y_j}$$

Now the LRT is defined as

$$\hat{\lambda}_n = \frac{L_n(\hat{\mu}_{MLE,\mathcal{H}_0}, \hat{\theta}_{MLE,\mathcal{H}_0})}{L_n(\hat{\mu}_{MLE,\Theta}, \hat{\theta}_{MLE,\Theta})}$$

Now, let $\hat{\mu}, \hat{\theta}$ to denote the MLE over the entire sample space Θ and $\hat{\mu}_0, \hat{\theta}_0$ denote the MLE over $\mathcal{H}_0$, we have

$$\hat{\lambda}_n = \frac{\hat{\theta}_0^{m+n} (\prod x_i)^{\hat{\theta}_0 - 1} (\prod y_j)^{\hat{\theta}_0 - 1}}{\hat{\mu}^n \hat{\theta}^m (\prod x_i)^{\hat{\mu} - 1} (\prod y_j)^{\hat{\theta} - 1}} = \left(\frac{\hat{\theta}_0}{\hat{\theta}}\right)^m \cdot \left(\frac{\hat{\theta}_0}{\hat{\mu}}\right)^n \cdot \left(\prod_{i=1}^n x_i\right)^{\hat{\theta}_0 - \hat{\mu}} \cdot \left(\prod_{j=1}^m y_j\right)^{\hat{\theta}_0 - \hat{\theta}}$$

by plugging in the corresponding values, we can easily conclude that (the last two terms simplify to 1):

$$\hat{\lambda}_n = \left(\frac{n+m}{m}\right)^m \cdot \left(\frac{n+m}{n}\right)^n \cdot (1-T)^m \cdot T^n$$

where $T = \dfrac{\sum \log x_i}{\sum \log x_i + \sum \log y_j}$. The critical region can be obtained by solving $\hat{\lambda}_n < C$ for some $C \in [0, 1]$.

Remark: Now note that $X_1, \cdots, X_n \sim Beta(\mu, 1)$, hence it is easy to see that $-\log X_i \sim Exponential(\mu)$ which has a pdf $\mu e^{-\mu x}$, similarly $-\log Y_j \sim Exponential(\theta)$. Using moment generating technique, we can also see that $-\sum \log X_i \sim Gamma(n, \mu)$ and $-\sum \log Y_j \sim Gamma(m, \theta)$, and **Basu's Theorem** tells us

$$T = \frac{\sum \log X_i}{\sum \log X_i + \log Y_j} \sim Beta(n, m).$$

3.4 Derivative Comes Again!

Let $X_1, \cdots, X_n$ be a random sample from a distribution with pdf $f(x,\theta) = \theta x^{\theta-1}$ for $0 \le x \le 1$, and $\theta > 0$ is the unknown parameter.

(i) Assume we wish to test $\mathcal{H}_0 : \theta = \theta_0$ v.s $\mathcal{H}_1 : \theta = \theta_1$ where $0 < \theta_0 < \theta_1$, use the NP lemma, show that the rejection region of the UMP test of level α takes the form

$$\sum_{i=1}^n \log X_i > k$$

for some k.

(ii) Get an expression of the value k in (i), under significance level α.

(iii) Now, we wish to test $\mathcal{H}_0 : \theta = \theta_0$ v.s $\mathcal{H}_1 : \theta \neq \theta_0$. Obtain a test statistic and get an approximate rejection region at $\alpha = 0.05$. Assume n is large.

Solution:

(i)

We use NP lemma to construct the ratio of the joint pdf:

$$\frac{p(\boldsymbol{x}, \theta_1)}{p(\boldsymbol{x}, \theta_0)} = \frac{\theta_1^n \cdot \prod_{i=1}^n x_i^{\theta_1 - 1}}{\theta_0^n \cdot \prod_{i=1}^n x_i^{\theta_0 - 1}} = \left(\frac{\theta_1}{\theta_0}\right)^n \cdot \prod_{i=1}^n x_i^{\theta_1 - \theta_0}$$

We know that the rejection region takes the form $\frac{p(\boldsymbol{x}, \theta_1)}{p(\boldsymbol{x}, \theta_0)} > k$ for some k. We take the logarithm of the expression above, and move the constant term to the other side of the equation, we will get

$$\sum_{i=1}^n \log X_i > k^*$$

for some constant k^*.

(ii)

Note that $\log X_i < 0$, so we aim to investigate the distribution of $Y_i = -\log X_i$. Note that

$$P(-\log X_i \le y) = P(X_i \ge e^{-y}) = \int_{e^{-y}}^1 \theta x^{\theta-1} dx$$

then the pdf of Y_i takes the form

$$f_{Y_i}(y) = \frac{d}{dy} \int_{e^{-y}}^1 \theta x^{\theta-1} dx = -\frac{d}{dy} \int_1^{e^{-y}} \theta x^{\theta-1} dx = -\theta (e^{-y})^{\theta-1} \cdot (-e^{-y}) = \theta e^{-\theta y}.$$

Hence we claim that $Y_i = -\log X_i \sim \textit{Exponential}(\theta)$, and then we can easily see that $-\sum_{i=1}^n \log X_i = \sum_{i=1}^n Y_i \sim \textit{Gamma}(n, \theta)$. So back to the rejection region in (i), we have

$$P\left(\sum_{i=1}^n \log X_i > k^*\right) = P\left(-k^* < -\sum_{i=1}^n \log X_i\right) = 1 - \alpha$$

Let $Y \sim \textit{Gamma}(n, \theta)$, then we know $P(Y > -k^*) = 1 - \alpha$, and recall the pdf of a Gamma distribution, we solve

$$\int_{-k^*}^{\infty} \frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x} dx = 1 - \alpha$$

where only $-k^*$ is unknown, so using a Gamma table we may compute the exact numerical value based on different α.

(iii)

To use the LR statistic, we first find the MLE of θ: Note that

$$\ell_n(\theta) = n \log \theta + (\theta - 1) \log \prod x_i \text{ , and } \frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{n}{\theta} + \log \prod x_i$$

and it can easily shown that the MLE of θ is given by

$$\hat{\theta}_{MLE} = -\frac{n}{\log \prod x_i}.$$

Then, we use the χ^2 approximation to find the rejection region. We have the approximate

$$\begin{aligned}
\mathcal{R} &= \left\{\boldsymbol{x} \in \mathcal{X} : -2\log[\lambda_n(\boldsymbol{X})] \geq \chi^2_{d,\alpha}\right\} \\
&= \left\{-2\log\left[\frac{L_n(\theta_0, \boldsymbol{x})}{L_n(\hat{\theta}_{MLE}, \boldsymbol{x})}\right] \geq \chi^2_{d,\alpha}\right\} \\
&= \left\{-2\log\left[\frac{\theta_0^n \prod x_i^{\theta_0 - 1}}{\hat{\theta}^n \prod x_i^{\hat{\theta}-1}}\right] \geq \chi^2_{d,\alpha}\right\} \\
&= \left\{-2\log \theta_0^n \prod x_i^{\theta_0 - 1} + 2 \log \hat{\theta}_n^n \prod x_i^{\hat{\theta}-1} \geq \chi^2_{d,\alpha}\right\}
\end{aligned}$$

In this example, $\alpha = 0.05, d = 1$, so we find $\chi^2_{1,0.05} = 3.841$. That is, we will reject $\mathcal{H}_0$ if

$$-2n \log \theta_0 - 2(\theta_0 - 1) \log \prod x_i + 2n \log \hat{\theta} + 2(\hat{\theta} - 1) \log \prod x_i \geq 3.841$$

where $\hat{\theta}$ is the MLE of θ.

4 Solved Extra Exercises in Confidence Interval