Statistics for Data Science

MSc Data Science WiSe 2019/20

Prof. Dr. Dirk Ostwald

Frequentist inference

(9) Confidence intervals

Bibliographic remarks

The presented material follows Held and Sabanés Bové (2014, Sections 3.2.2 - 3.2.3) for the definitions of confidence intervals and pivots, Moeschlin (2001, Sections 4.1 - 4.5) for the exact confidence interval examples, and Wasserman (2004, Sections 6.3.2 and 9.7) for the interpretation of confidence intervals and the notion of approximate confidence intervals, respectively.

Confidence intervals

- Definition, interpretation, and pivots
- Z, U, T, and Wald statistics
- Exact confidence intervals
- Approximate confidence intervals for MLEs

Confidence intervals

- **Definition, interpretation, and pivots**
- Exemplary exact confidence intervals
- Exemplary approximate confidence intervals

Definition (δ-confidence interval (δ-CI))

Given a parametric statistical model $\mathcal{P}$ with PMF or PDF p_θ, a sample $X := X_1, ..., X_n \sim p_\theta$, an $\delta \in]0,1[$, and two statistics $b_l(X)$ and $b_u(X)$, a δ-*confidence interval* is an interval $C_n = [b_l, b_u]$, such that

$$\mathbb{P}_\theta\left(\theta \in C_n\right) = \mathbb{P}_\theta\left(b_l(X) \leq \theta \leq b_u(X)\right) = \delta \text{ for all } \theta \in \Theta. \tag{1}$$

δ is called the *confidence level* or *coverage probability*. The random variables $b_l(X)$ and $b_u(X)$ are the lower and upper bounds of the confidence interval.

Remarks

- C_n is a random interval, because $b_l(X)$ and $b_u(X)$ are random variables.
- θ is fixed (not random) and unknown.
- A δ-confidence interval covers the true θ with probability δ.
- Often $\delta = 0.95$, resulting in 95%-*confidence intervals*.
- Confidence intervals with $b_l = -\infty$ or $b_u = \infty$ are called *one-sided*.

Two interpretations of δ confidence intervals

(1) If an experiment is repeated over and over, the δ-confidence interval will contain the true, but unknown, parameter in $\delta \cdot 100\%$ of all cases. More technically, for repeated random samples form a distribution with unknown parameter θ, a δ- confidence interval will cover θ in $\delta \cdot 100\%$ of all cases.

(2) Consider a sequence of experiments with unrelated parameters $\theta_1, \theta_2, ...$ and imagine constructing δ-confidence intervals for the sequence of unrelated parameters $\theta_1, \theta_2, ...$ Then $\delta \cdot 100\%$ of the confidence intervals will contain the true, but unknown, parameter value.

Definition (Exact and approximate Pivots)

An *exact pivot* is a function of the data $X_1, ..., X_n$ and the true, but unknown, parameter θ with distribution not depending on θ. An *approximate pivot* is a pivot whose distribution does not asymptotically depend on the true parameter θ.

Remarks

- A pivot is a statistic depending on the true θ, with distribution independent of θ.
- Pivots can be used to construct confidence intervals valid for all values of θ.
- Pivots must not depend on nuisance parameters η.
- For nuisance parameters η, the pivot's distribution must not depend on θ nor η.
- CIs based on exact pivots will be referred to as *exact CIs*.
- CIs based on approximate pivots will be referred to as *approximate CIs*.
- Typical exact and approximate pivots are the T and Wald statistics, respectively.

Confidence intervals

- Definition, interpretation, and pivots
- **Z, U, T, and Wald statistics**
- Exact confidence intervals
- Approximate confidence intervals for MLEs

Theorem (Z statistic)

Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ be a normally distributed random sample and let

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2}$$

denote the sample mean. Then the *Z statistic* is defined as

$$Z := \frac{\sqrt{n}}{\sigma} \left(\bar{X}_n - \mu \right) \tag{3}$$

and is distributed according to standard normal distribution,

$$Z \sim N(0, 1), \tag{4}$$

such that a PDF for X is given by

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2 \right). \tag{5}$$

Proof

- Casella and Berger (2002, p. 218 - 219)

Theorem (U statistic)

Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ be a normally distributed random sample and let

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \tag{6}$$

denote the sample mean and sample variance, respectively. Then the *U statistic* is defined as

$$U := \frac{n-1}{\sigma^2} S^2 \tag{7}$$

and is distributed according to a chi-squared distribution with $n-1$ degrees of freedom,

$$U \sim \chi^2(n-1), \tag{8}$$

such that a PDF for U is given by

$$p_{n-1}(u) = \frac{1}{\Gamma((n-1)/2)2^{(n-1)/2}} u^{((n-1)/2)-1} \exp\left(-\frac{1}{2}u\right). \tag{9}$$

Proof

- Casella and Berger (2002, p. 219)

Theorem (T statistic)

Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ be a normally distributed random sample and let

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S_n := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} \tag{10}$$

denote the sample mean and sample standard deviation, respectively. Then the T *statistic* is defined as

$$T := \sqrt{n}\frac{\bar{X} - \mu}{S} \tag{11}$$

and is distributed according to a t-distribution with $n-1$ degrees of freedom,

$$T \sim \mathrm{T}(n-1), \tag{12}$$

such that a PDF for T is given by

$$p_{n-1}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{((n-1)\pi)^{1/2}} \left(\left(1 + \frac{t^2}{n-1}\right)^{n/2}\right)^{-1}. \tag{13}$$

Proof

- Casella and Berger (2002, p. 219)

Theorem (Wald statistic)

Let $X_1, ..., X_n \sim p_\theta$ be a random sample, let $\hat{\theta}_n^{ML}$ denote the maximum likelihood estimator for θ, and let

$$I_n\left(\hat{\theta}_n^{ML}\right) := \frac{d^2}{d\theta^2}\ell_n\left(\hat{\theta}_n^{ML}\right) \tag{14}$$

denote the observed Fisher information of the random sample. Then the *Wald statistic* is defined as

$$W_n := \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}\left(\hat{\theta}_n^{ML} - \theta\right) \tag{15}$$

and is asymptotically distributed according to a standard normal distribution,

$$W_n \overset{a}{\sim} N(0,1), \tag{16}$$

such that an approximate PDF for W_n is given by

$$p(w_n) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}w_n^2\right). \tag{17}$$

Proof

- Held and Sabanés Bové (2014, Section 4.2.3, 4.2.4), also see Lecture (8).

Confidence intervals

- Definition, interpretation, and pivots
- Z, U, T, and Wald statistics
- **Exemplary exact confidence intervals**
- Exemplary approximate confidence intervals

The typical construction of confidence intervals commonly involves

1. The definition of the parametric statistical model.
2. The definition of a statistic and the assessment of its distribution.
3. The verification of the confidence condition.
4. The formulation of the confidence interval.

In the following, we demonstrate the above for exact confidence intervals of

1. The expectation of a normal distribution with known variance.
2. The expectation of a normal distribution with unknown variance.
3. The variance of a normal distribution.

Example (Expectation of a normal distribution, variance known)

1. Parametric statistical model

   Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ denote a random sample of a parametric statistical model with known variance parameter $\sigma^2 > 0$ and unknown expectation parameter $\mu \in \mathbb{R}$. We develop a δ-confidence interval for μ.

2. Statistic and its distribution

   We consider the statistic

$$Z := \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu). \tag{18}$$

   We have seen previously, that $Z \sim N(0, 1)$. Note that Z is a function of $X_1, ..., X_n$ (by means of $\bar{X}$) and μ, while its distribution does not depend on μ. Z is thus an exact pivot.

Example (Expectation of a normal distribution, variance known)

3. Confidence condition

For $\delta \in]0,1[$, let $z_1 := \Phi^{-1}\left(\frac{1-\delta}{2}\right)$ and $z_2 := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$ denote the respective percentiles of $N(0,1)$. For example, for $\delta = 0.95$, $z_1 = \Phi^{-1}(0.025) = -1.96$ and $z_2 = \Phi^{-1}(0.975) = 1.96$. Note that with the symmetry of $N(0,1)$, $z_1 = -z_2$. Then

$$\mathbb{P}\left(-z_2 \leq Z \leq z_2\right) = \delta. \tag{19}$$

Thus, also

$$\begin{aligned}
\mathbb{P}\left(-z_2 \leq \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \leq z_2\right) &= \mathbb{P}\left(-\frac{\sigma}{\sqrt{n}}z_2 \leq \bar{X} - \mu \leq \frac{\sigma}{\sqrt{n}}z_2\right) \\
&= \mathbb{P}\left(-\bar{X} - \frac{\sigma}{\sqrt{n}}z_2 \leq -\mu \leq -\bar{X} + \frac{\sigma}{\sqrt{n}}z_2\right) \\
&= \mathbb{P}\left(\bar{X} + \frac{\sigma}{\sqrt{n}}z_2 \geq \mu \geq \bar{X} - \frac{\sigma}{\sqrt{n}}z_2\right) \\
&= \mathbb{P}\left(\bar{X} + \frac{\sigma}{\sqrt{n}}z_2 \geq \mu \geq \bar{X} - \frac{\sigma}{\sqrt{n}}z_2\right) \\
&= \mathbb{P}\left(\bar{X} - \frac{\sigma}{\sqrt{n}}z_2 \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}}z_2\right) \\
&= \delta.
\end{aligned} \tag{20}$$

Example (Expectation of a normal distribution, variance known)

4. Formulation of the confidence interval

   For a random sample $X_1, ..., X_n \sim N(\mu, \sigma^2)$ with known variance parameter σ^2 and unknown expectation parameter μ, $\alpha \in]0,1[$, $\delta := \delta$, and $z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$, set

$$C_n := \left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X} + \frac{\sigma}{\sqrt{n}} z_\delta\right]. \tag{21}$$

   Then

$$\mathbb{P}(\mu \in C_n) = \delta. \tag{22}$$

   and C_n is a δ-confidence interval for μ. Note that because $\bar{X}$ is a random variable, C_n is random.

□

Example (Expectation of a normal distribution, variance unknown)

1. Parametric statistical model

   Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ denote a random sample from a parametric statistical model with unknown expectation parameter μ and unknown variance parameter $\sigma^2 > 0$. We develop a δ-confidence interval for μ.

2. Statistic and its distribution

   We consider the statistic

$$T := \frac{\sqrt{n}}{S}(\bar{X} - a), \text{ where } S := \sqrt{\frac{1}{n-1}(X_i - \bar{X})^2}. \tag{23}$$

   We have seen previously, that $T \sim T(n-1)$. Note that T is a function of $X_1, ..., X_n$ (by means of $\bar{X}$ and S), while its distribution does not depend on neither μ and σ^2. Also note that σ^2 here attains the role of a nuisance parameter, and neither T nor its distribution depend on σ^2.

Example (Expectation of a normal distribution, variance unknown)

1. Confidence condition

   For $\delta \in]0,1[$, let $t_1 := \Psi_{n-1}^{-1}\left(\frac{1-\delta}{2}\right)$ and $t_2 := \Psi_{n-1}^{-1}\left(\frac{1+\delta}{2}\right)$ denote the respective percentiles of $t(n-1)$. For example, for $n = 10$ and $\delta = 0.95$, $t_1 = \Psi_9^{-1}(0.025) = -2.26$ and $t_2 = \Psi_9^{-1}(0.975) = 2.26$. Note that with the symmetry of $t(n-1)$, $t_1 = -t_2$. Then

$$\mathbb{P}\left(-t_2 \leq T \leq t_2\right) = \delta. \tag{24}$$

   Thus, also

$$\begin{aligned}
\mathbb{P}\left(-t_2 \leq \frac{\sqrt{n}}{S}(\bar{X} - \mu) \leq t_2\right) &= \mathbb{P}\left(-\frac{S}{\sqrt{n}}t_2 \leq \bar{X} - \mu \leq \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(-\bar{X} - \frac{S}{\sqrt{n}}t_2 \leq -\mu \leq -\bar{X} + \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(\bar{X} + \frac{S}{\sqrt{n}}t_2 \geq \mu \geq \bar{X} - \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(\bar{X} + \frac{S}{\sqrt{n}}t_2 \geq \mu \geq \bar{X} - \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(\bar{X} - \frac{S}{\sqrt{n}}t_2 \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}}t_2\right) \\
&= \delta.
\end{aligned} \tag{25}$$

Example (Expectation of a normal distribution, variance unknown)

4. Formulation of the confidence interval

   For a random sample $X_1, ..., X_n \sim N(\mu, \sigma^2)$ with unknown expectation parameter μ and unknown variance parameter μ, $\delta \in]0,1[$, and $t_\delta := \Psi^{-1}_{n-1}\left(\frac{1+\delta}{2}\right)$, set

$$C_n := \left[\bar{X} - \frac{S}{\sqrt{n}} t_\delta, \bar{X} + \frac{S}{\sqrt{n}} t_\delta\right]. \tag{26}$$

   Then

$$\mathbb{P}(\mu \in C_n) = \delta. \tag{27}$$

   and C_n is a δ-confidence interval for μ. Note that because $\bar{X}$ and S are random variables, C_n is random.

□

Example (Variance of a normal distribution)

1. Parametric statistical model

   Let $X_1, ..., X_n \sim N(\mu, \sigma^2)$ denote a random sample of a parametric statistical model with unknown variance parameter $\sigma^2 > 0$ and known or unknown expectation parameter $\mu \in \mathbb{R}$. We develop a δ-confidence interval for σ^2.

2. Statistic and its distribution

   We consider the statistic

$$U := \frac{n-1}{\sigma^2} S^2, \text{ where } S^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 . \tag{28}$$

   We have seen previously, that $U \sim \chi^2(n-1)$. Note that U is a function of $X_1, ..., X_n$ (by means of S^2) and σ^2, while its distribution does not depend on σ^2. U is thus an exact pivot.

Example (Variance of a normal distribution)

1. Confidence condition

For $\delta \in]0,1[$, let $\xi_1 := \Xi^{-1}_{n-1}\left(\frac{1-\delta}{2}\right)$ and $\xi_2 := \Xi^{-1}_{n-1}\left(\frac{1+\delta}{2}\right)$ denote the respective percentiles of $\chi^2(n-1)$. For example, for for $n = 10$ and $\delta = 0.95$, $\xi_1 := \Xi^{-1}_{9}(0.025) = 2.70$ and $\xi_2 := \Xi^{-1}_{9}(0.975) = 19.0$. Then

$$\mathbb{P}\left(\xi_1 \le U \le \xi_2\right) = \delta. \tag{29}$$

Thus, also

$$\begin{aligned}\mathbb{P}\left(\xi_1 \le \frac{n-1}{\sigma^2}S^2 \le \xi_2\right) &= \mathbb{P}\left(\xi_1^{-1} \ge \frac{\sigma^2}{(n-1)S^2} \ge \xi_2^{-1}\right)\\ &= \mathbb{P}\left(\frac{(n-1)S^2}{\xi_1} \ge \sigma^2 \ge \frac{(n-1)S^2}{\xi_2}\right)\\ &= \mathbb{P}\left(\frac{(n-1)S^2}{\xi_2} \le \sigma^2 \le \frac{(n-1)S^2)}{\xi_1}\right)\\ &= \delta.\end{aligned} \tag{30}$$

Example (Variance of a normal distribution)

4. Formulation of the confidence interval

   For a random sample $X_1, ..., X_n \sim N(\mu, \sigma^2)$ with unknown expectation parameter μ and unknown variance parameter σ^2, $\delta \in]0,1[$, $\xi_1 := \Xi_{n-1}^{-1}\left(\frac{1-\delta}{2}\right)$ and $\xi_2 := \Xi_{n-1}^{-1}\left(\frac{1+\delta}{2}\right)$, set

$$C_n := \left[\frac{(n-1)S^2}{\xi_2}, \frac{(n-1)S^2)}{\xi_1}\right]. \tag{31}$$

   Then

$$\mathbb{P}(\sigma^2 \in C_n) = \delta. \tag{32}$$

   and C_n is a δ-confidence interval for σ^2. Note that because S^2 is a random variables, C_n is random.

□

Confidence intervals

- Definition, interpretation, and pivots
- Z, U, T, and Wald statistics
- Exemplary exact confidence intervals
- **Exemplary approximate confidence intervals**

The typical construction of confidence intervals commonly involves

1. The definition of the parametric statistical model.
2. The definition of a statistic and the assessment of its distribution.
3. The verification of the confidence condition.
4. The formulation of the confidence interval.

We next demonstrate the above for approximate confidence intervals of

1. Maximum likelihood parameter estimators.
2. The MLE of a Bernoulli distribution parameter.

Example (Maximum likelihood estimators)

1. Parametric statistical model

   Let $X_1, ..., X_n \sim p_\theta$ denote a random sample of a parametric statistical model. We develop a δ-confidence interval for θ.

2. Statistic and its distribution

   We consider the Wald statistic

$$W_n := \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}\left(\hat{\theta}_n^{ML} - \theta\right) \tag{33}$$

   We have seen previously, that $W_n \overset{a}{\sim} N(0,1)$. Note that W_n is a function of $X_1, ..., X_n$ and θ, while its asymptotic distribution does not depend on θ. W_n is thus an approximate pivot.

Example (Maximum likelihood estimators)

3. Confidence condition

For $\delta \in]0,1[$, let $z_1 := \Phi^{-1}\left(\frac{1-\delta}{2}\right)$ and $z_2 := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$ denote the respective percentiles of $N(0,1)$. For example, for $\delta = 0.95$, $z_1 = \Phi^{-1}(0.025) = -1.96$ and $z_2 = \Phi^{-1}(0.975) = 1.96$. Note that with the symmetry of $N(0,1)$, $z_1 = -z_2$. Then

$$\mathbb{P}\left(-z_2 \leq W_n \leq z_2\right) \to \delta \text{ for } n \to \infty. \tag{34}$$

Thus, also

$$\begin{aligned}
&\mathbb{P}\left(-z_2 \leq \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}\left(\hat{\theta}_n^{ML} - \theta\right) \leq z_2\right) \\
&= \mathbb{P}\left(-\sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2 \leq \hat{\theta}_n^{ML} - \theta \leq \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2\right) \\
&= \mathbb{P}\left(-\hat{\theta}_n^{ML} - \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2 \leq -\theta \leq -\hat{\theta}_n^{ML} + \frac{1}{SE} z_2\right) \\
&= \mathbb{P}\left(\hat{\theta}_n^{ML} + \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2 \geq \theta \geq \hat{\theta}_n^{ML} - \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2\right) \\
&= \mathbb{P}\left(\hat{\theta}_n^{ML} - \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2 \leq \mu \leq \hat{\theta}_n^{ML} + \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_2\right) \\
&\to \delta \text{ for } n \to \infty.
\end{aligned} \tag{35}$$

Example (Maximum likelihood estimators)

4. Formulation of the confidence interval

For a random sample $X_1, ..., X_n \sim p_\theta$, $\delta \in]0,1[$, and $z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$, set

$$C_n := \left[\hat{\theta}_n^{ML} - \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_\delta, \hat{\theta}_n^{ML} + \sqrt{I_n\left(\hat{\theta}_n^{ML}\right)}^{-1} z_\delta\right]. \tag{36}$$

Then

$$\mathbb{P}(\theta \in C_n) \to \delta \text{ for } n \to \infty. \tag{37}$$

and C_n is an approximate δ-confidence interval for θ. Note that because $\hat{\theta}_n^{ML}$ is a random variable, C_n is random.

□

Example (Expectation of a Bernoulli distribution)

Let $X_1, ..., X_n \sim \text{Bern}(\mu)$ denote a Bernoulli distributed random sample of size n and let

$$\hat{\mu}_n^{ML} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{38}$$

denote the maximum likelihood estimator of μ. Then an approximate 95%-confidence interval for μ is given by

$$C_n = \left[\hat{\mu}_n^{ML} - 1.96 \sqrt{\frac{\hat{\mu}_n^{ML} \left(1 - \hat{\mu}_n^{ML}\right)}{n}}, \hat{\mu}_n^{ML} + 1.96 \sqrt{\frac{\hat{\mu}_n^{ML} \left(1 - \hat{\mu}_n^{ML}\right)}{n}} \right], \tag{39}$$

because $z_{0.95} \approx 1.96$ and

$$I_n \left(\hat{\mu}_n^{ML} \right) = \frac{n}{\hat{\mu}_n^{ML} \left(1 - \hat{\mu}_n^{ML}\right)}. \tag{40}$$

Example (Expectation of a Bernoulli distribution)

Proof Let $X \sim \text{Bern}(\mu)$. Then the Fisher information of the random variable X is given by

$$
\begin{aligned}
I(\mu) &:= -\frac{d^2}{d\mu^2}\ell_1(\mu) \\
&= -\frac{d^2}{d\mu^2} \ln p_\mu(x) \\
&= -\frac{d^2}{d\mu^2} \left(x \ln \mu + (1-x)\ln(1-\mu)\right) \\
&= -\frac{d}{d\mu} \left(\frac{d}{d\mu} \left(x \ln \mu + (1-x)\ln(1-\mu)\right) \right) \\
&= -\frac{d}{d\mu} \left(\frac{x}{\mu} + \frac{(1-x)}{1-\mu} \right) \\
&= -\left(-\frac{x}{\mu^2} - \frac{(1-x)^{2}}{1-\mu} \right) \\
&= \frac{x}{\mu^2} + \frac{(1-x)^{2}}{1-\mu}
\end{aligned}
\tag{41}
$$

Example (Expectation of a Bernoulli distribution)

Proof

Furthermore, the expected Fisher information of the random variable X is given by

$$\begin{aligned} J(\mu) &= \mathbb{E}_\mu(I(\mu)) \\ &= \mathbb{E}_\mu\left(\frac{X}{\mu^2} + \frac{(1-X)}{1-\mu}^2\right) \\ &= \frac{\mathbb{E}_\mu(X)}{\mu^2} + \frac{(1-\mathbb{E}_\mu(X))}{1-\mu}^2 \\ &= \frac{\mu}{\mu^2} + \frac{(1-\mu)}{1-\mu}^2 \\ &= \frac{1}{\mu(1-\mu)} \end{aligned} \tag{42}$$

With the additivity property of the expected Fisher information and the definition of the observed Fisher information, it then follows immediately, that

$$J_n(\mu) = \frac{n}{\mu(1-\mu)} \tag{43}$$

and

$$J_n\left(\hat{\mu}_n^{ML}\right) = \frac{n}{\hat{\mu}_n^{ML}\left(1-\hat{\mu}_n^{ML}\right)}, \tag{44}$$

respectively.

□

References

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference*, volume 10. Springer.

Moeschlin, O. (2001). *Angewandte Statistik*. FernUniversitaet in Hagen.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer.