Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v17n3p572

Comparing Fisher Information in Continuous, Dichotomized, and Discretized Data: A Pedagogical Perspective with Illustrations
By Bartolucci and Pennoni

15 December 2024

Electronic Journal of Applied Statistical Analysis
Vol. 17, Issue 03, December 2024, 572-585
DOI: 10.1285/i20705948v17n3p572

Comparing Fisher Information in Continuous, Dichotomized, and Discretized Data: A Pedagogical Perspective with Illustrations

Francesco Bartolucci[a] and Fulvia Pennoni*[b]

[a] *University of Perugia, Department of Economics, Via Pascoli, 20 - 06123 Perugia, Italy*
[b] *University of Milano-Bicocca, Department of Statistics and Quantitative Methods, Via Bicocca degli Arcimboldi, 8 - 20126 Milan, Italy*

15 December 2024

In this pedagogical paper we provide an illustration of Fisher information through examples based on the dichotomization and discretization of continuous variables following a Gaussian distribution. In these examples, we demonstrate the information loss and quantify the increase in sample size necessary to recover this loss. To facilitate understanding and application, we also provide R codes for performing illustrative simulations. This approach aims to improve the teaching and learning experience of Fisher information in statistical education.

keywords: Binary data, Cramér-Rao inequality, Fisher information, Maximum likelihood estimation, Ordinal data

MSC: 62A01, 6204, 97M10

1 Introduction

In mathematical statistics, information is a crucial concept that dates back to Sir Ronald Aylmer Fisher (1890-1962), who discussed it in several works; see Fisher (1922) and Fisher (1925) for early papers. Sir Francis Ysidro Edgeworth (1845-1926) was the first to introduce the concept of information; see Edgeworth (1908) and the review provided

*Corresponding author: fulvia.pennoni@unimib.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

in Savage (1976). The Fisher information is illustrated in most statistical textbooks, such as Cox and Hinkley (1979), Lehmann and Casella (1998), and Casella and Berger (2002); just to mention a recent overview, see Ly et al. (2017). A gentle video tutorial is provided at `https://www.youtube.com/watch?v=pneluWj-U-o`. As stated in Frieden (2004, p. 23): "*Fisher information is a simple and intuitive concept. [...] once understood, the concept gives strong payoff - one might call it "phenomen-all"- in scope of application. It's simply worth learning.*".

With reference to a single random variable following a parametric model $X \sim f_\theta(x)$ that depends on a single parameter θ, the Fisher information is defined as

$$\bar{\mathrm{I}}(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right] = -\mathrm{E}_\theta\left[\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)\right],$$

where usual regularity conditions are assumed. Referred to a sample of n independent observations from the assumed model, the overall Fisher's information simply becomes $\mathrm{I}(\theta) = n\bar{\mathrm{I}}(\theta)$. These definitions may be extended to the multiparametric case in a natural way, obtaining average and overall information matrices.

In teaching Fisher information, a typical question from students is why this name has been historically adopted. Generally, the answer that $\bar{\mathrm{I}}(\theta)$ is a measure of the amount of information on θ provided by a single observation and $\mathrm{I}(\theta)$ is that provided by a sample of size n, without specific motivations, is considered unsatisfactory. Classic examples related to the estimation of the mean under the Gaussian (with known variance), Bernoulli, and Poisson models illustrate this point. As is well known, for these models we have the following results:

$$X \sim \mathrm{N}(\mu, \sigma^2): \qquad \bar{\mathrm{I}}(\mu) = 1/\sigma^2, \tag{1}$$

$$X \sim \mathrm{Bern}(p): \qquad \bar{\mathrm{I}}(p) = 1/[p(1-p)], \tag{2}$$

$$X \sim \mathrm{Pois}(\lambda): \qquad \bar{\mathrm{I}}(\lambda) = 1/\lambda. \tag{3}$$

In these examples, the Fisher information is equal to the reciprocal of the population variance. Therefore, when the variance is equal under all three models, there is the same information on the population mean. This seems to contradict the general belief that binary or count data have a smaller information content than continuous data; see, among others, Gertsbakh (1995) and Royston et al. (2006).

A related definition is that of the Cramér-Rao inequality, first introduced in Fréchet (1943) and extended by Rao (1945) and Cramér (1946); see Savage (1972) for historical details. According to this inequality, for an unbiased estimator the lower bound of the variance (and then of its mean squared error) is $1/\mathrm{I}(\theta)$. The larger is $\mathrm{I}(\theta)$, the smaller is the lower bound of variance of the estimator, and so the greater its precision when the variance touches this lower bound. Therefore, $\mathrm{I}(\theta)$ is a measure of efficiency of the estimation procedure. Moreover, $1/\bar{\mathrm{I}}(\theta)$ is the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta)$, where $\hat{\theta}$ is the maximum likelihood estimator (MLE) of θ, so that its variance may be approximated with $1/\mathrm{I}(\theta)$. The Cramér-Rao inequality can be adopted as a paradigm to explain the Fisher information, although for the three examples in (1), (2), and (3),

we obtain the same lower bound for the same sample size and population variance; this lower bound is reached by the sample mean in all these examples.

To provide adequate explanations in teaching Fisher information, while retaining simplicity, in this note we propose two examples based on the comparison between the information contained in continuous data that are directly observable and that contained in the corresponding dichotomized or discretized data. Note that comparisons between these two situations are already available in the literature even in more sophisticated contexts; see, for instance, Shentu and Xie (2010). However, our focus here is purely on information rather than on other aspects and on presenting results from a teaching perspective. In this regard, we also provide in the appendix some R codes (R Core Team, 2023) to perform illustrative simulations. For related studies from a different perspective, we refer the reader to works such as Park (1996), where a plot is suggested to visually represent the Fisher information for any set of consecutive order statistics in a parametric distribution; see also Tsairidis et al. (2001), Park and Balakrishnan (2009) and Park et al. (2011), where the properties of Fisher information are explored under various no censoring conditions: quantal, complete, incomplete, and hybrid censoring schemes; finally, see Barakat et al. (2021) and Husseiny et al. (2022), where other features of the Fisher information concerning order statistics are explored.

The remainder of this note is organized as follows. In Section 2 we deal with the case of dichotomizing a continuous variable so as to obtain a binary variable. The case of discretizing a continuous variable in an arbitrary number of categories is dealt with in Section 3. Section 4 provides some final conclusions, while the illustrative R codes are reported in Appendix.

2 Observing continuous versus dichotomized data

For a Gaussian model with known variance, $X \sim \mathrm{N}(\mu, \sigma^2)$, the Fisher information on μ is $\bar{\mathrm{I}}_X(\mu) = 1/\sigma^2$, as noted previously. Under this model, the MLE of μ, which is the sample mean $\hat{\mu}_1 = \bar{X}$, has variance $V_1(\hat{\mu}_1) = \sigma^2/n$. The asymptotic (in this case, also finite-sample) variance of $\sqrt{n}(\hat{\mu}_1 - \mu)$ is $\bar{V}_1(\hat{\mu}_1) = \sigma^2$.

Now consider a dichotomized version of X resulting in $Y = 1(X > \tau)$, where τ is a known cutpoint and $1(\cdot)$ denotes the indicator function. Obviously, Y has Bernoulli distribution, in symbols $Y \sim \mathrm{Bern}(p)$, with success probability

$$p = 1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian distribution, which can also be expressed as

$$p = \Phi(\delta), \quad \delta = \frac{\mu - \tau}{\sigma}.$$

Note that

$$\frac{\partial p}{\partial \mu} = \frac{\phi(\delta)}{\sigma},$$

where $\phi(\cdot)$ is the density function of the standard Gaussian distribution, and

$$\frac{\partial^2 p}{\partial \mu^2} = -\frac{\phi(\delta)\delta}{\sigma^2}.$$

The log-likelihood function for Y is given by

$$\log f_\mu(y) = y \log(p) + (1-y)\log(1-p),$$

and the score function (for a single observation) is

$$\frac{\partial \log f_\mu(y)}{\partial \mu} = \left(\frac{y}{p} - \frac{1-y}{1-p}\right)\frac{\partial p}{\partial \mu}, \tag{4}$$

and

$$\frac{\partial^2 \log f_\mu(y)}{\partial \mu^2} = -\left(\frac{y}{p^2} + \frac{1-y}{(1-p)^2}\right)\left(\frac{\partial p}{\partial \mu}\right)^2 + \left(\frac{y}{p} - \frac{1-y}{1-p}\right)\frac{\partial^2 p}{\partial \mu^2},$$

so that

$$\bar{\mathrm{I}}_Y(\mu) = \frac{1}{p(1-p)}\left(\frac{\partial p}{\partial \mu}\right)^2 = \frac{\phi^2(\delta)}{\sigma^2 p(1-p)}.$$

Also note that, from (4), the MLE of μ based on a sample of binary observations drawn from the model for Y is obtained by solving the equation

$$\bar{Y} - \Phi\left(\frac{\mu - \tau}{\sigma}\right) = 0,$$

leading to the solution

$$\hat{\mu}_2 = \tau + \sigma\Phi^{-1}(\bar{Y}), \tag{5}$$

where $\bar{Y}$ is the sample mean. The asymptotic variance of $\sqrt{n}(\hat{\mu}_2 - \mu)$ is

$$\bar{V}(\hat{\mu}_2) = \frac{\sigma^2 p(1-p)}{\Phi^2(\delta)}.$$

We can now compare the information on μ based on two samples, one drawn from the distribution of X and the other from the distribution of Y. In particular, we consider the following ratio between the average Fisher information functions:

$$\frac{\bar{\mathrm{I}}_Y(\mu)}{\bar{\mathrm{I}}_X(\mu)} = \frac{\phi^2(\delta)}{p(1-p)} = R(\delta).$$

Note that $R(\delta)$ is also equal to the ratio between the asymptotic variances $\bar{V}(\hat{\mu}_1)$ and $\bar{V}(\hat{\mu}_2)$.

Function $R(\delta)$, which is symmetric, is represented in Figure 1, while values of this function for certain values of δ are reported in Table 1. These results can be obtained by the R code reported in Appendix A1, where we also demonstrate an illustrative simulation. It is worth noting that $R(\delta)$ is always smaller than 1, and its maximum value

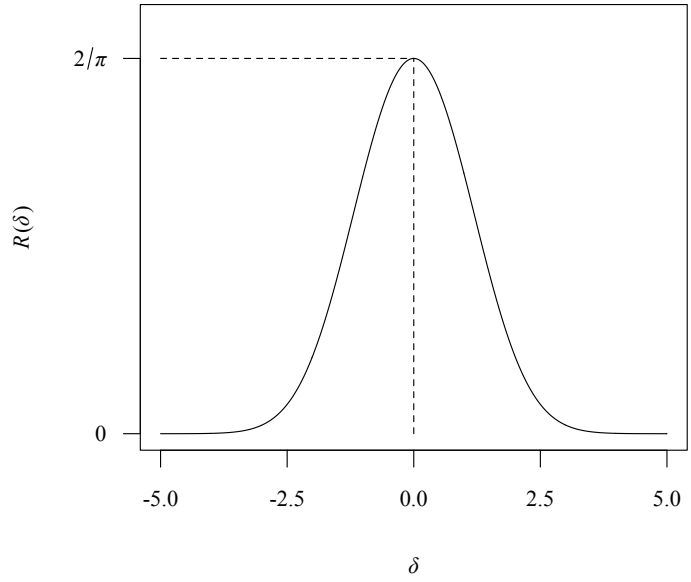

Figure 1: *Representation of $R(\delta)$ with respect to δ.*

is reached for $\delta = 0$, when it is equal to $R(0) = 2/\pi \approx 0.637$; therefore, this maximum value is attained when $\mu = \tau$ regardless of σ^2. Moreover, the loss of information increases as δ deviates from 0; that is, as the distance between μ and τ increases (with σ^2 constant) or as σ^2 decreases (with $\mu - \tau$ constant). Table 1 also presents values of $100/R(\delta)$, where this ratio can be interpreted as the sample size of dichotomized variables having the same information content as a sample of size 100 of continuous outcomes.

3 Observing continuous versus discretized data

The example proposed in the previous section can be extended to the case of continuous data that are discretized into a finite number k of categories, labelled from 0 to $k-1$. This approach is commonly adopted in the analysis of ordinal data; see, among others, McCullagh (1980) and Agresti (2013, Ch. 8). This discretization is based on the ordered cutpoints $\tau_1 < \cdots < \tau_{k-1}$ such that

$$Y = y \quad \text{if} \quad \tau_y \leq x < \tau_{y+1}, \quad y = 0, \ldots, k-1,$$

where $\tau_0 = -\infty$ and $\tau_k = +\infty$ for completeness.

We also introduce the indicator variables $Z_y = 1(Y = y)$ and the random vector $\boldsymbol{Z} = (Z_0, \ldots, Z_{k-1})'$, which is equivalent to Y. Under the assumption of a Gaussian

δ	$R(\delta)$	$100/R(\delta)$
-5.00	0.00000771	12968690.3
-2.50	0.04978708	2008.6
-1.00	0.43862886	228.0
-0.50	0.58099165	172.1
-0.25	0.62230110	160.7
-0.10	0.63431001	157.7
0.00	0.63661977	157.1
0.10	0.63431001	157.7
0.25	0.62230110	160.7
1.00	0.43862886	228.0
2.50	0.04978708	2008.6
5.00	0.00000771	12968690.3

Table 1: *Values of* $R(\delta)$ *and* $100/R(\delta)$ *for certain typical values of* δ.

distribution for X, as formulated at the beginning of Section 2, $\boldsymbol{Z}$ follows a generalized Bernoulli distribution with probability mass function

$$f_\mu(\boldsymbol{z}) = \prod_{y=0}^{k-1} p_y^{z_y},$$

where z_y is a realization of Z_y and these realizations are collected in $\boldsymbol{z}$, and

$$p_y = \begin{cases} 1 - \Phi(\delta_1) & \text{if } y = 0, \\ \Phi(\delta_y) - \Phi(\delta_{y+1}) & \text{if } y = 1, \ldots, k-2, \\ \Phi(\delta_{k-1}) & \text{if } y = k-1, \end{cases}$$

with $\delta_y = (\mu - \tau_y)/\sigma$; consequently we have

$$\frac{\partial p_y}{\partial \mu} = \frac{1}{\sigma} \begin{cases} -\phi(\delta_1) & \text{if } y = 0, \\ \phi(\delta_y) - \phi(\delta_{y+1}) & \text{if } y = 1, \ldots, k-2, \\ \phi(\delta_{k-1}) & \text{if } y = k-1, \end{cases}$$

and

$$\frac{\partial^2 p_y}{\partial \mu^2} = -\frac{1}{\sigma^2} \begin{cases} -\phi(\delta_1)\delta_1 & \text{if } y = 0, \\ \phi(\delta_y)\delta_y - \phi(\delta_{y+1})\delta_{y+1} & \text{if } y = 1, \ldots, k-2, \\ \phi(\delta_{k-1})\delta_{k-1} & \text{if } y = k-1. \end{cases}$$

Now note that

$$\log f_\mu(\boldsymbol{z}) = \sum_{y=0}^{k-1} z_y \log p_y,$$

so that

$$\frac{\partial \log f_\mu(\boldsymbol{z})}{\partial \mu} = \sum_{y=0}^{k-1} \frac{z_y}{p_y} \frac{\partial p_y}{\partial \mu},$$

and

$$\frac{\partial^2 \log f_\mu(\boldsymbol{z})}{\partial \mu^2} = \sum_{y=0}^{k-1} \left[\frac{z_y}{p_y} \frac{\partial^2 p_y}{\partial \mu^2} - \frac{z_y}{p_y^2} \left(\frac{\partial p_y}{\partial \mu} \right)^2 \right],$$

leading to the Fisher information

$$\bar{\mathrm{I}}_{\boldsymbol{Z}}(\mu) = \sum_{y=0}^{k-1} \frac{1}{p_y} \left(\frac{\partial p_y}{\partial \mu} \right)^2 = \frac{1}{\sigma^2} \left\{ \frac{\phi^2(\delta_1)}{p_0} + \sum_{y=1}^{k-2} \frac{[\phi(\delta_y) - \Phi(\delta_{y+1})]^2}{p_y} + \frac{\phi^2(\delta_{k-1})}{p_{k-1}} \right\}.$$

In this case, for the MLE of μ denoted by $\hat{\mu}_3$, there is not an explicit formula as that in (5), but an iterative algorithm, such as the Newton-Raphson algorithm, is required. As an illustrative example, this algorithm is implemented in the R function provided in Appendix A2.

In order to demonstrate the loss of information due to the discretization, which is now quantified by the ratio

$$R(\boldsymbol{\delta}) = \frac{\bar{\mathrm{I}}_{\boldsymbol{Z}}(\mu)}{\bar{\mathrm{I}}_X(\mu)} = \frac{\phi^2(\delta_1)}{p_0} + \sum_{y=1}^{k-2} \frac{[\phi(\delta_y) - \phi(\delta_{y+1})]^2}{p_y} + \frac{\phi^2(\delta_{k-1})}{p_{k-1}},$$

we consider three different scenarios with k categories equal to 3, 5, and 10, and cutpoints $\tau_j = \alpha + \beta \xi_j$, where $\xi_j = 2(j-1)/(k-2) - 1$, $j = 1, \ldots, k-1$. These points are equally spaced between -1 and 1, while $\mu = 0$ and $\sigma^2 = 1$ in all cases. Results for different values of α and β are reported in Table 2.

Appendix A3 presents the R code that can be used to perform an illustrative simulation for computing $R(\boldsymbol{\delta})$.

4 Conclusion

The Fisher information is a foundational concept in statistical inference across various disciplines. By quantifying the amount of information provided by data about model parameters, the Fisher information enhances our understanding of the reliability and efficiency of statistical estimators. For example, in vaccine trials, this concept aids in evaluating in which measure a sample of a certain size can provide reliable estimates of vaccine effectiveness. In radar signal processing, understanding Fisher information is crucial for accurately estimating parameters such as target location and velocity from noisy measurements. In psychometrics and education, it helps in understanding how effectively a test discriminates between different levels of ability and how sample size affects the precision of ability estimates.

Dichotomization of continuous variables is prevalent in clinical studies and other fields. For instance, in drug discovery, chemical potencies are often dichotomized during data

k	α	β	$R(\boldsymbol{\delta})$	$100/R(\boldsymbol{\delta})$
3	0	1	0.73807618	135.5
3	0	2	0.25626440	390.2
3	1	1	0.69576936	143.7
3	1	2	0.45076098	221.8
3	5	1	0.00056554	176823.5
3	5	2	0.01456986	6863.5
5	0	1	0.91205868	109.6
5	0	2	0.86971179	115.0
5	1	1	0.79858440	125.2
5	1	2	0.85707561	116.7
5	5	1	0.00056621	176612.9
5	5	2	0.01458162	6857.9
10	0	1	0.93327913	107.1
10	0	2	0.97526729	102.5
10	1	1	0.81323549	123.0
10	1	2	0.95110931	105.1
10	5	1	0.00056685	176412.2
10	5	2	0.01464070	6830.3

Table 2: *Values of* $R(\boldsymbol{\delta})$ *and* $100/R(\boldsymbol{\delta})$ *for certain values of* δ*; these values do not depend on the sign of* α*.*

analysis. In medical research, clinical measurements like blood pressure have conventional thresholds used for diagnoses. In social sciences and psychology, categorized outcomes are common, with binary response variables frequently modeled using logistic and probit regression models. The conclusions drawn from the previous examples align with the general principle that discretized data contain less information than the original continuous data. When continuous variables are dichotomized, as illustrated in Section 2, the most favorable scenario occurs when $\delta = 0$, where δ measures the relative distance (in terms of standard deviation) between the population mean and the cutpoint. According to the results in Table 1, in this case, the amount of information in a sample of continuous data of size n is comparable to that contained in a sample of dichotomized data of size $1.571n$. However, when $\delta \neq 0$ the loss of information could be much more severe. For instance, the size of the dichotomized sample would need to increase to $2.28n$ for $|\delta| = 1$ and to $20.09n$ for $|\delta| = 2.5$.

However, when continuous variables are discretized in k categories, several factors must be considered. From the results in Table 2, the primary conclusion is that the severity of the information loss tends to decrease as k increases, assuming other factors

remain constant. For instance, with $\alpha = 0$ and $\beta = 2$, the sample size required for discretized variables is $3.90n$ with three categories and $1.03n$ with ten categories. Here, α and β define the sequence of cutpoints. This is reasonable because a finer discretization corresponds to more informative categories. However, even with many categories, such as ten, the necessary increase in sample size to compensate for discretization could be substantial. This occurs, for example, when μ lies outside the range of the cutpoints $\tau_1, \ldots, \tau_{k-1}$, as observed when $\alpha = 5$ and $\beta = 1$.

Finally note that the examples proposed in Sections 2 and 3 may be extended to regression models that include covariates. In these models, instead of observing a continuous response variable, we may observe its dichotomized or discretized versions. This comparison can involve linear regression models against probit or ordinal probit models. For further insights, we refer the reader to Winship and Mare (1984). Furthermore, it has been demonstrated that analysing rather than the available continuous variable its discretized version may offer certain advantages, such as enhancing robustness against specific types of data contamination. For detailed discussions, see Shentu and Xie (2010).

Appendix

A1. R code for the case of dichotomized variables: Simulations

The following R code can be used to compute simulated and theoretical values of function $R(\delta)$, which represent the ratio between the average Fisher information functions when continuous variables are dichotomized, as outlined in Section 2.

- Simulation setup: Theoretical values

```
mu = 1; si2 = 4
tau = -0.5
n = 1000
```

- Perform simulations

```
si = sqrt(si2)
mu1v = mu2v = rep(0,10000)
for(it in 1:10000){
  xv = rnorm(n,mu,si)
  yv = 1*(xv>tau)
  mu1 = mean(xv)
  mu2 = tau+si*qnorm(mean(yv))
  mu1v[it] = mu1; mu2v[it] = mu2
}
```

- Print output

– Simulated value of $R(\delta)$

```
print(var(mu1v)/var(mu2v))
```

```
#> [1] 0.525802
```

– Theoretical value of $R(\delta)$

```
de = (mu-tau)/si
print(dnorm(de)^2/(pnorm(de)*(1-pnorm(de))))
```

```
#> [1] 0.5174023
```

A2. R code for the case of discretized variables: Maximum likelihood estimation of μ

In the following we show the code of the functions used to obtain the MLE of μ as described Section 3 based on the Newton-Raphson algorithm. We also report some results when running these functions.

- Main function named nr_mu.R

```
nr_mu = function(yv,tauv,si2){
# preliminaries
  n = length(yv)
  k = max(yv)+1
  si = sqrt(si2)

# compute probabilities and their derivatives
  dev = (mu-tauv)/si
  pv = 1 - pnorm(dev[1])
  for(y in 1:(k-2)) pv = c(pv,pnorm(dev[y])-pnorm(dev[y+1]))
  pv = c(pv,pnorm(dev[k-1]))
  d1v = -dnorm(dev[1])/si
  for(y in 1:(k-2)) d1v = c(d1v,dnorm(dev[y])/si-dnorm(dev[y+1])/si)
  d1v = c(d1v,dnorm(dev[k-1])/si)
  d2v = dnorm(dev[1])*dev[1]/si2
  for(y in 1:(k-2)) d2v = c(d2v,-(dnorm(dev[y])*dev[y]/si2-
                                  dnorm(dev[y+1])*dev[y+1]/si2))
  d2v = c(d2v, - dnorm(dev[k-1])*dev[k-1]/si2)

# log-likelihood, score and observed information
  lk = sc = J = 0
  for(i in 1:n){
    lk = lk+log(pv[yv[i]+1])
```

```
    sc = sc+d1v[yv[i]+1]/pv[yv[i]+1]
    J = J-(d2v[yv[i]+1]/pv[yv[i]+1]-d1v[yv[i]+1]^2/pv[yv[i]+1]^2)
  }
# iterate until convergence
  # print(c(0,mu,lk))
  lko = lk; it = 0
  while((lk-lko)/abs(lko)>10^-10 | it==0){
    lko = lk; it = it+1
# update parameter
    mu = mu+sc/J
# compute probabilities and their derivatives
    dev = (mu-tauv)/si
    pv = 1 - pnorm(dev[1])
    for(y in 1:(k-2)) pv = c(pv,pnorm(dev[y])-pnorm(dev[y+1]))
    pv = c(pv,pnorm(dev[k-1]))
    d1v = - dnorm(dev[1])/si
    for(y in 1:(k-2)) d1v = c(d1v,dnorm(dev[y])/si-dnorm(dev[y+1])/si)
    d1v = c(d1v,dnorm(dev[k-1])/si)
    d2v = dnorm(dev[1])*dev[1]/si2
    for(y in 1:(k-2)) d2v = c(d2v,-(dnorm(dev[y])*dev[y]/si2-
                                    dnorm(dev[y+1])*dev[y+1]/si2))
    d2v = c(d2v,-dnorm(dev[k-1])*dev[k-1]/si2)

# log-likelihood, score and observed information
    lk = sc = J = 0
    for(i in 1:n){
      lk = lk+log(pv[yv[i]+1])
      sc = sc+d1v[yv[i]+1]/pv[yv[i]+1]
      J = J-(d2v[yv[i]+1]/pv[yv[i]+1]-d1v[yv[i]+1]^2/pv[yv[i]+1]^2)
    }
# print output
    # print(c(it,mu,lk,lk-lko))
  }

# output
  out = list(mu=mu,lk=lk,sc=sc,J=J,pv=pv)
}
```

A3. R code for the case of discretized variables: Simulations

In the following we show the code to perform the simulation described in Section 3.

- Simulation setup

```
k = 3
mu = 0; si2 = 1
al = 0; be = 1
n = 1000
```

- Compute cutpoints

```
si = sqrt(si2)
si = sqrt(si2)
xiv = 2*((1:(k-1))-1)/(k-2)-1
tauv = al+be*xiv
```

- Perform simulations

```
si = sqrt(si2)
mu1v = mu3v = rep(0,1000)
for(it in 1:1000){
  xv = rnorm(n,mu,si)
  yv = rep(0,n)
  for(y in 1:(k-1)) yv = yv+(xv>tauv[y])
  out = nr_mu(yv,tauv,si2)
  mu1v[it] = mean(xv)
  mu3v[it] = out$mu
}
```

- Print output
  - Simulated value of $R(\boldsymbol{\delta})$

```
print(var(mu1v)/var(mu3v))
```

```
#> [1] 0.7788731
```

  - Theoretical value of $R(\boldsymbol{\delta})$

```
dev = (mu-tauv)/si
pv = 1-pnorm(dev[1])
for(y in 1:(k-2)) pv = c(pv,pnorm(dev[y])-pnorm(dev[y+1]))
pv = c(pv,pnorm(dev[k-1]))
R = dnorm(dev[1])^2/pv[1]
for(y in 1:(k-2)) R = R+(dnorm(dev[y])-dnorm(dev[y+1]))^2/pv[y+1]
R = R+dnorm(dev[k-1])^2/pv[k]
print(R)
```

```
#> [1] 0.7380762
```

References

Agresti, A. (2013). *Categorical Data Analysis, 3nd Edition*. John Wiley & Sons, Hoboken, New Jersey.

Barakat, H. M., Nigm, E. M., and Husseiny, I. A. (2021). Measures of information in order statistics and their concomitants for the single iterated Farlie–Gumbel–Morgenstern bivariate distribution. *Mathematical Population Studies*, 28:154–175.

Casella, G. and Berger, R. L. (2002). *Statistical Inference, 2nd Edition*. Duxbury Press, Pacific Grove.

Cox, D. R. and Hinkley, D. V. (1979). *Theoretical Statistics*. CRC Press, New York.

Cramér, H. (1946). A contribution to the theory of statistical estimation. *Scandinavian Actuarial Journal*, 29:85–94.

Edgeworth, F. Y. (1908). On the probable errors of frequency-constants. *Journal of the Royal Statistical Society*, 71:381–397.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, 222:309–368.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725.

Fréchet, M. (1943). Sur l'extension de certaines evaluations statistiques de petits echantillons. *Revue Internationale de Statistique*, 11:182–205.

Frieden, B. R. (2004). *Science from Fisher Information*. Cambridge University Press, Cambridge.

Gertsbakh, I. (1995). On the Fisher information in type-I censored and quantal response data. *Statistics & Probability Letters*, 23:297–306.

Husseiny, I., Alawady, M., Barakat, H., and Abd Elgawad, M. (2022). Information measures for order statistics and their concomitants from cambanis bivariate family. *Communications in Statistics - Theory and Methods*, 8:1–17.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation, 2nd Edition*. Springer Verlag, New York.

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P., and Wagenmakers, E.-J. (2017). A tutorial on Fisher information. *Journal of Mathematical Psychology*, 80:40–55.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42:109–142.

Park, S. (1996). Fisher information in order statistics. *Journal of the American Statistical Association*, 91:385–390.

Park, S. and Balakrishnan, N. (2009). On simple calculation of the fisher information in hybrid censoring schemes. *Statistics & Probability Letters*, 79:1311–1319.

Park, S., Balakrishnan, N., and Kim, S. W. (2011). Fisher information in progressive hybrid censoring schemes. *Statistics*, 45:623–631.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91.

Royston, P., Altman, D. G., and Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25:127–141.

Savage, L. J. (1972). *The Foundation of Statistics*. Wiley, 2nd ed., New York.

Savage, L. J. (1976). On rereading R. A. Fisher (with discussion). *Annals of Statistics*, 29:441–500.

Shentu, Y. and Xie, M. (2010). A note on dichotomization of continuous response variable in the presence of contamination and model misspecification. *Statistics in Medicine*, 29:2200–2214.

Tsairidis, C., Zografos, K., Ferentinos, K., and Papaioannou, T. (2001). Information in quantal response data and random censoring. *Annals of the Institute of Statistical Mathematics*, 53:528–542.

Winship, C. and Mare, R. (1984). Regression models for ordinal variables. *American Sociological Review*, 1:512–525.