# Notes: Statistics

Matthew Low

*Last Updated: August 7, 2023*

.

## Contents

# 1 Integrals, Moments, and Asymmetries

We review some relevant distributions and calculations needed for parameter estimation.

### Data

Consider a dataset $\{x_k\}$ where $k = 1, \ldots, N$. We have

$$\overline{x_k} = \frac{1}{N}\sum_{k=1}^{N} x_k, \tag{1}$$

$$\mathrm{Var}(x_k) = \left(\frac{1}{N}\sum_{k=1}^{N} x_k^2\right) - \overline{x_k}^2, \tag{2}$$

$$\mathrm{Var}(\overline{x_k}) = \left(\frac{1}{N^2}\sum_{k=1}^{N} x_k^2\right) - \frac{1}{N}\overline{x_k}^2. \tag{3}$$

If the dataset of $N$ values is split into two samples, called $N^+$ and $N^-$ (where $N^+ + N^- = N$) then we have a binomially distributed variable. We call the asymmetry $A$ of $\{x_k\}$ to be

$$A(x_k) = \frac{N^+ - N^-}{N^+ + N^-}. \tag{4}$$

We can normalize such that $p = N^+/N$ and $1 - p = N^-/N$. We then have

$$A(x_k) = p - (1 - p) = 2p - 1, \tag{5}$$

This means that the asymmetry inherits its variance from the variance of $p$

$$\mathrm{Var}(A) = 4\ \mathrm{Var}(p). \tag{6}$$

Similarily

$$\mathrm{Var}(p) = \frac{1}{N^2}\ \mathrm{Var}(N^+). \tag{7}$$

Since $N^+$ is binomially distributed we have

$$\mathrm{Var}(N^+) = Np(1-p) = \frac{N^+N^-}{N}. \tag{8}$$

The $p$ version is useful for calculating from the known distributions while the $N^+$ version is useful for evaluating in a dataset. Thus we find

$$\mathrm{Var}(p) = \frac{p(1-p)}{N} = \frac{N^+N^-}{N^3}, \qquad \mathrm{Var}(A) = \frac{4p(1-p)}{N} = \frac{4N^+N^-}{N^3}. \tag{9}$$

## Example: Spin Correlation

Consider the following distribution

$$\frac{1}{\sigma}\frac{d\sigma}{dx_{ij}} = \frac{1}{2}\left(C_{ij}x_{ij} - 1\right)\log|x_{ij}|. \tag{10}$$

Assuming the data is well described by this distribution, we can extract parameters by computing the moments of this distribution. We first ensure that it is normalized

$$\int_{-1}^{1}\frac{1}{\sigma}\frac{d\sigma}{dx_{ij}}dx_{ij} = \int_{-1}^{1}\frac{1}{2}\left(C_{ij}x_{ij} - 1\right)\log|x_{ij}|dx_{ij} = \int_{0}^{1}(-1)\log x_{ij}dx_{ij} = 1. \tag{11}$$

The normalization is independent of $C_{ij}$ since the $C_{ij}$ component is odd and just introduces an asymmetry. The mean of this distribution is

$$\begin{aligned} \overline{x_{ij}} = \langle x_{ij}\rangle &= \int_{-1}^{1} x_{ij}\frac{1}{\sigma}\frac{d\sigma}{dx_{ij}}dx_{ij} = \int_{-1}^{1}\frac{x_{ij}}{2}\left(C_{ij}x_{ij} - 1\right)\log|x_{ij}|dx_{ij} \\ &= C_{ij}\int_{0}^{1} x_{ij}^2\log x_{ij}dx_{ij} = -\frac{C_{ij}}{9}. \end{aligned} \tag{12}$$

The variance of this distribution is

$$\begin{aligned} \mathrm{Var}(x_{ij}) &= \langle (x_{ij} - \overline{x_{ij}})^2\rangle = \langle x_{ij}^2\rangle - \overline{x_{ij}}^2 \\ &= \left(\int_{-1}^{1}\frac{x_{ij}^2}{2}\left(C_{ij}x_{ij} - 1\right)\log|x_{ij}|dx_{ij}\right) - \overline{x_{ij}}^2 \\ &= \left(\int_{0}^{1}(-x_{ij}^2)\log x_{ij}dx_{ij}\right) - \overline{x_{ij}}^2 = \frac{1}{9} - \overline{x_{ij}}^2. \end{aligned} \tag{13}$$

It can also be useful to express $\overline{x_{ij}}^2 = C_{ij}^2/81$ which emphasizes that for $\mathcal{O}(1)$ values of $C_{ij}$ the variance is quite close to $1/9$. The variance on the mean is suppressed by a factor of $1/N$ where $N$ is the number of data points

$$\mathrm{Var}(\overline{x_{ij}}) = \frac{1}{N}\mathrm{Var}(x_{ij}) = \frac{1}{N}\left(\frac{1}{9} - \overline{x_{ij}^2}\right). \tag{14}$$

The parameter $C_{ij}$ can then be extracted as

$$\boxed{C_{ij} = -9\overline{x_{ij}} \pm \sqrt{\frac{9 - 81\overline{x_{ij}}^2}{N}}.} \tag{15}$$

Equivalently we would say that $\sigma(C_{ij}) \approx 3/\sqrt{N}$.

Next, using the asymmetry we have

$$A = \frac{C_{ij}}{4} \tag{16}$$

and with a variance of

$$\mathrm{Var}(A) = \frac{1}{N}\left(1 - \frac{C_{ij}^2}{16}\right). \tag{17}$$

$$\boxed{C_{ij} = 4A(x_{ij}) \pm \frac{4}{\sqrt{N}}\sqrt{1 - \frac{C_{ij}^2}{16}}.} \tag{18}$$

## Example: Spin Analyzing Power

Another useful distribution is

$$\frac{1}{\Gamma}\frac{d\Gamma}{dx_i} = \frac{1}{2}\left(1 + \kappa_i x_i\right). \tag{19}$$

For this distribution we find

$$\overline{x_i} = \frac{\kappa_i}{3}, \tag{20}$$

$$\mathrm{Var}(x_i) = \frac{1}{3} - \left(\frac{\kappa_i}{3}\right)^2, \tag{21}$$

so that we have

$$\boxed{\kappa_i = 3\overline{x_i} \pm \sqrt{\frac{3 - 9\overline{x_i}^2}{N}}.} \tag{22}$$