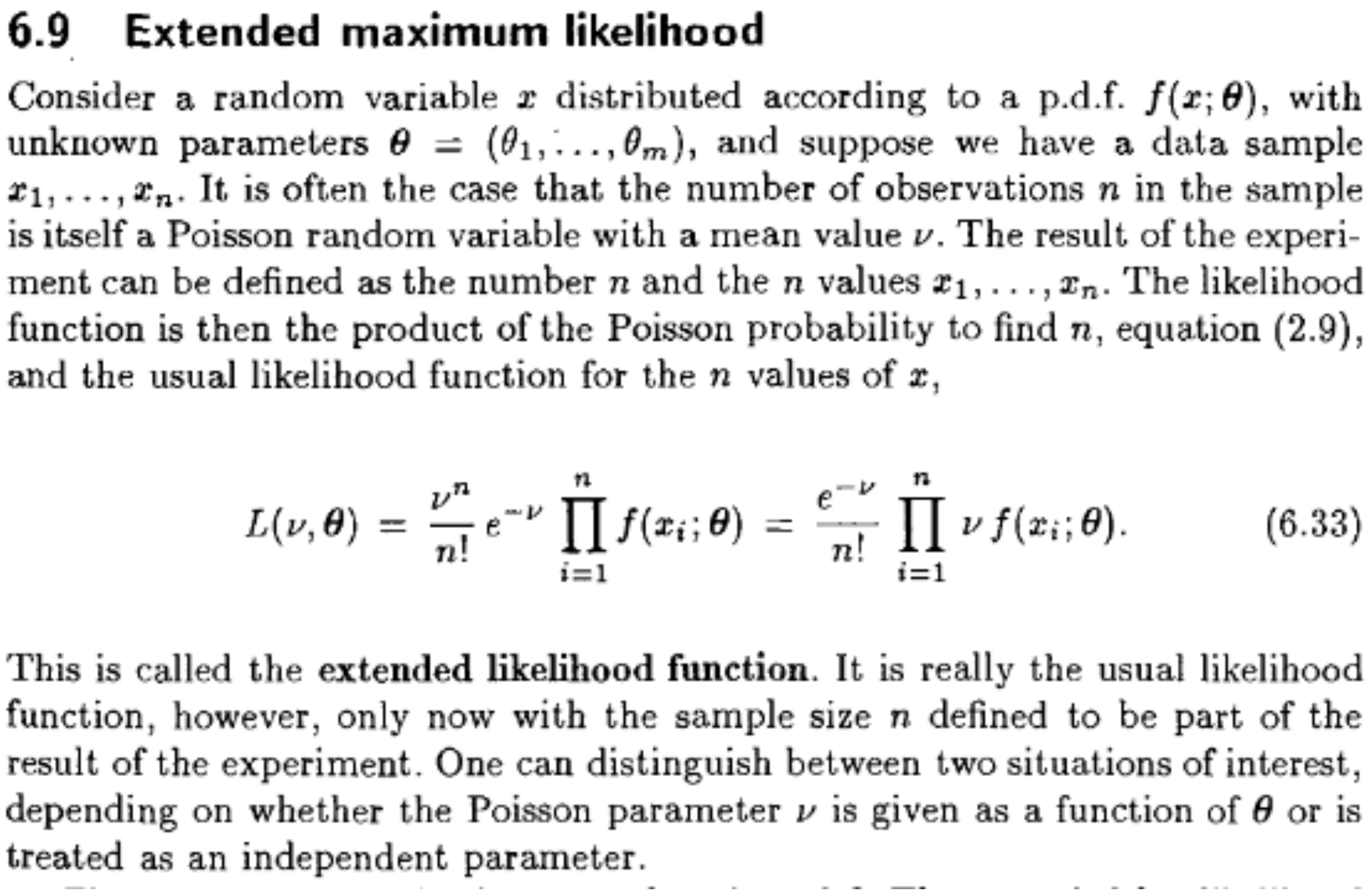
Esercitazione-3a
approfondimento/addendum

Laboratorio Analisi Dati / Alexis Pompili



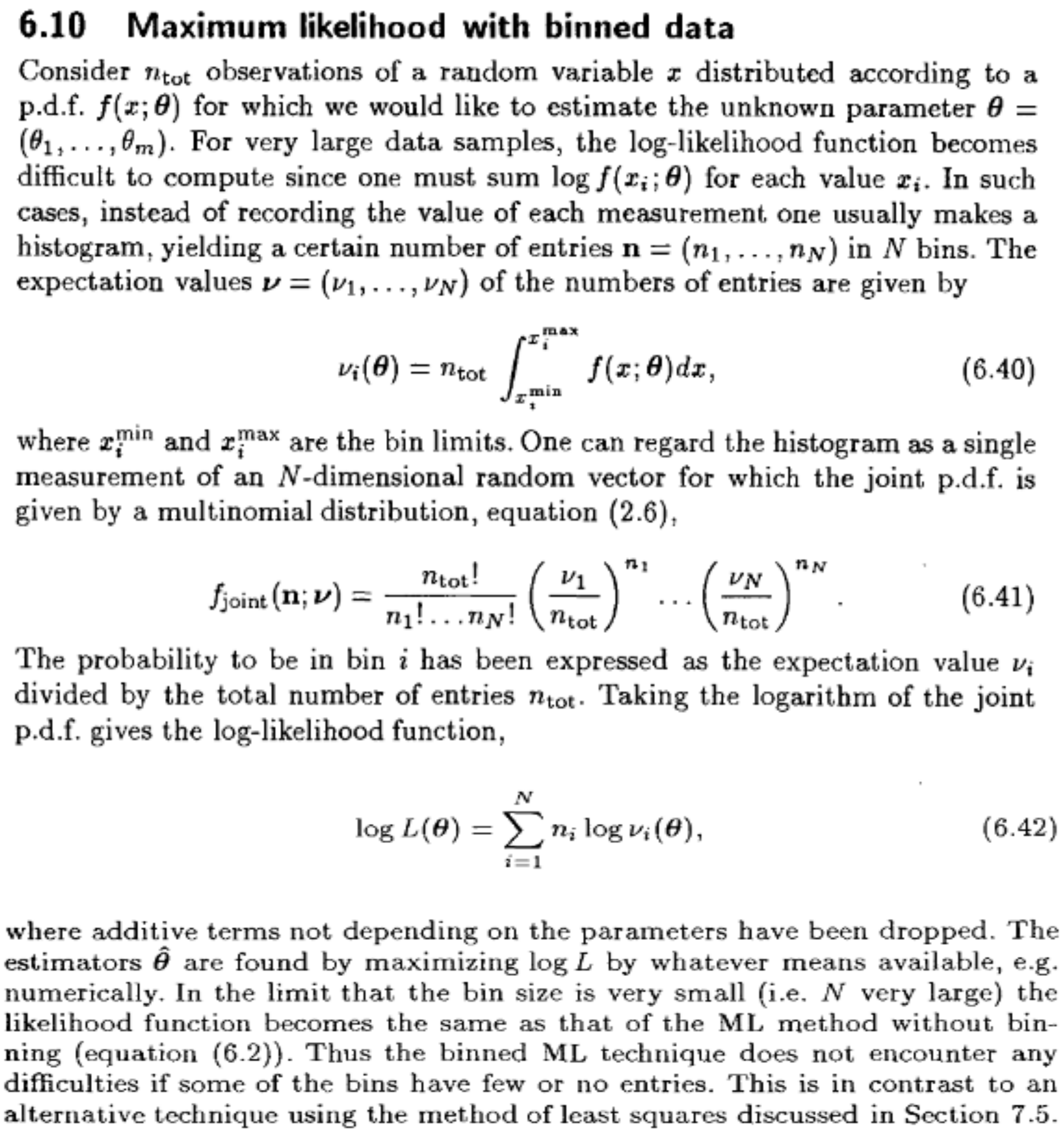
6.9 Extended maximum likelihood

Consider a random variable x distributed according to a p.d.f. $f(x;\boldsymbol{\theta})$, with unknown parameters $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)$, and suppose we have a data sample $x_1,\ldots,x_n$. It is often the case that the number of observations n in the sample is itself a Poisson random variable with a mean value ν. The result of the experiment can be defined as the number n and the n values $x_1,\ldots,x_n$. The likelihood function is then the product of the Poisson probability to find n, equation (2.9), and the usual likelihood function for the n values of x,

$$L(\nu,\boldsymbol{\theta}) = \frac{\nu^n}{n!}e^{-\nu}\prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}) = \frac{e^{-\nu}}{n!}\prod_{i=1}^{n}\nu f(x_i;\boldsymbol{\theta}). \qquad (6.33)$$

This is called the **extended likelihood function**. It is really the usual likelihood function, however, only now with the sample size n defined to be part of the result of the experiment. One can distinguish between two situations of interest, depending on whether the Poisson parameter ν is given as a function of $\boldsymbol{\theta}$ or is treated as an independent parameter.

Source: **G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998**



6.10 Maximum likelihood with binned data

Consider n_{tot} observations of a random variable x distributed according to a p.d.f. $f(x;\boldsymbol{\theta})$ for which we would like to estimate the unknown parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. For very large data samples, the log-likelihood function becomes difficult to compute since one must sum $\log f(x_i;\boldsymbol{\theta})$ for each value x_i. In such cases, instead of recording the value of each measurement one usually makes a histogram, yielding a certain number of entries $\mathbf{n} = (n_1, \ldots, n_N)$ in N bins. The expectation values $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ of the numbers of entries are given by

$$\nu_i(\boldsymbol{\theta}) = n_{\mathrm{tot}} \int_{x_i^{\mathrm{min}}}^{x_i^{\mathrm{max}}} f(x;\boldsymbol{\theta})dx, \tag{6.40}$$

where x_i^{min} and x_i^{max} are the bin limits. One can regard the histogram as a single measurement of an N-dimensional random vector for which the joint p.d.f. is given by a multinomial distribution, equation (2.6),

$$f_{\mathrm{joint}}(\mathbf{n};\boldsymbol{\nu}) = \frac{n_{\mathrm{tot}}!}{n_1! \ldots n_N!} \left(\frac{\nu_1}{n_{\mathrm{tot}}}\right)^{n_1} \cdots \left(\frac{\nu_N}{n_{\mathrm{tot}}}\right)^{n_N}. \tag{6.41}$$

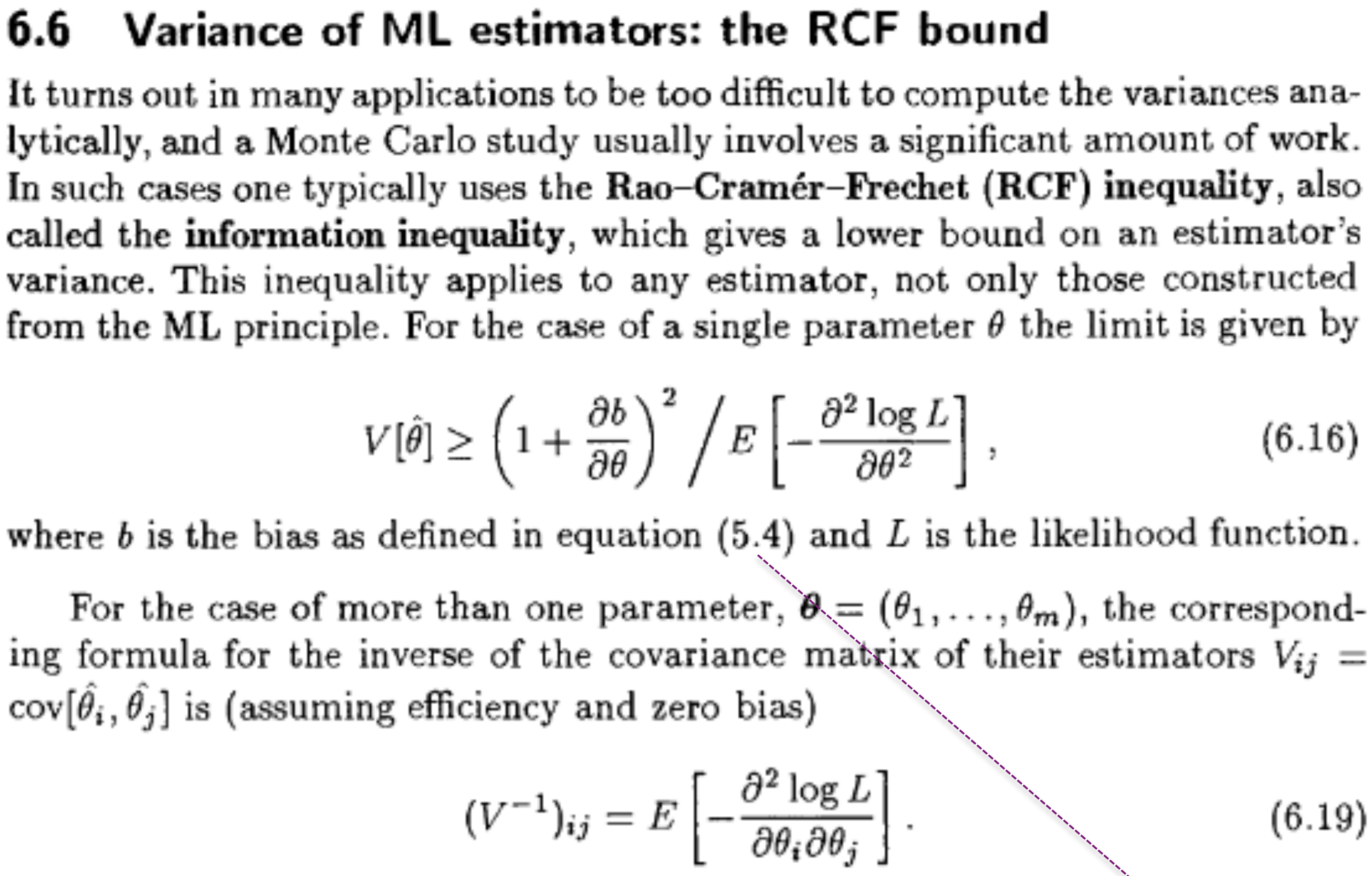
The probability to be in bin i has been expressed as the expectation value ν_i divided by the total number of entries n_{tot}. Taking the logarithm of the joint p.d.f. gives the log-likelihood function,

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \log \nu_i(\boldsymbol{\theta}), \tag{6.42}$$

where additive terms not depending on the parameters have been dropped. The estimators $\hat{\boldsymbol{\theta}}$ are found by maximizing $\log L$ by whatever means available, e.g. numerically. In the limit that the bin size is very small (i.e. N very large) the likelihood function becomes the same as that of the ML method without binning (equation (6.2)). Thus the binned ML technique does not encounter any difficulties if some of the bins have few or no entries. This is in contrast to an alternative technique using the method of least squares discussed in Section 7.5.

Source: **G.Cowan, Statistical Data Analysis, Clarendon Press Oxford, 1998**

Come fa MINUIT a calcolare la matrice di covarianza?



6.6 Variance of ML estimators: the RCF bound

It turns out in many applications to be too difficult to compute the variances analytically, and a Monte Carlo study usually involves a significant amount of work. In such cases one typically uses the **Rao–Cramér–Frechet (RCF) inequality**, also called the **information inequality**, which gives a lower bound on an estimator's variance. This inequality applies to any estimator, not only those constructed from the ML principle. For the case of a single parameter θ the limit is given by

$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 \Big/ E\left[-\frac{\partial^2 \log L}{\partial \theta^2}\right] , \qquad (6.16)$$

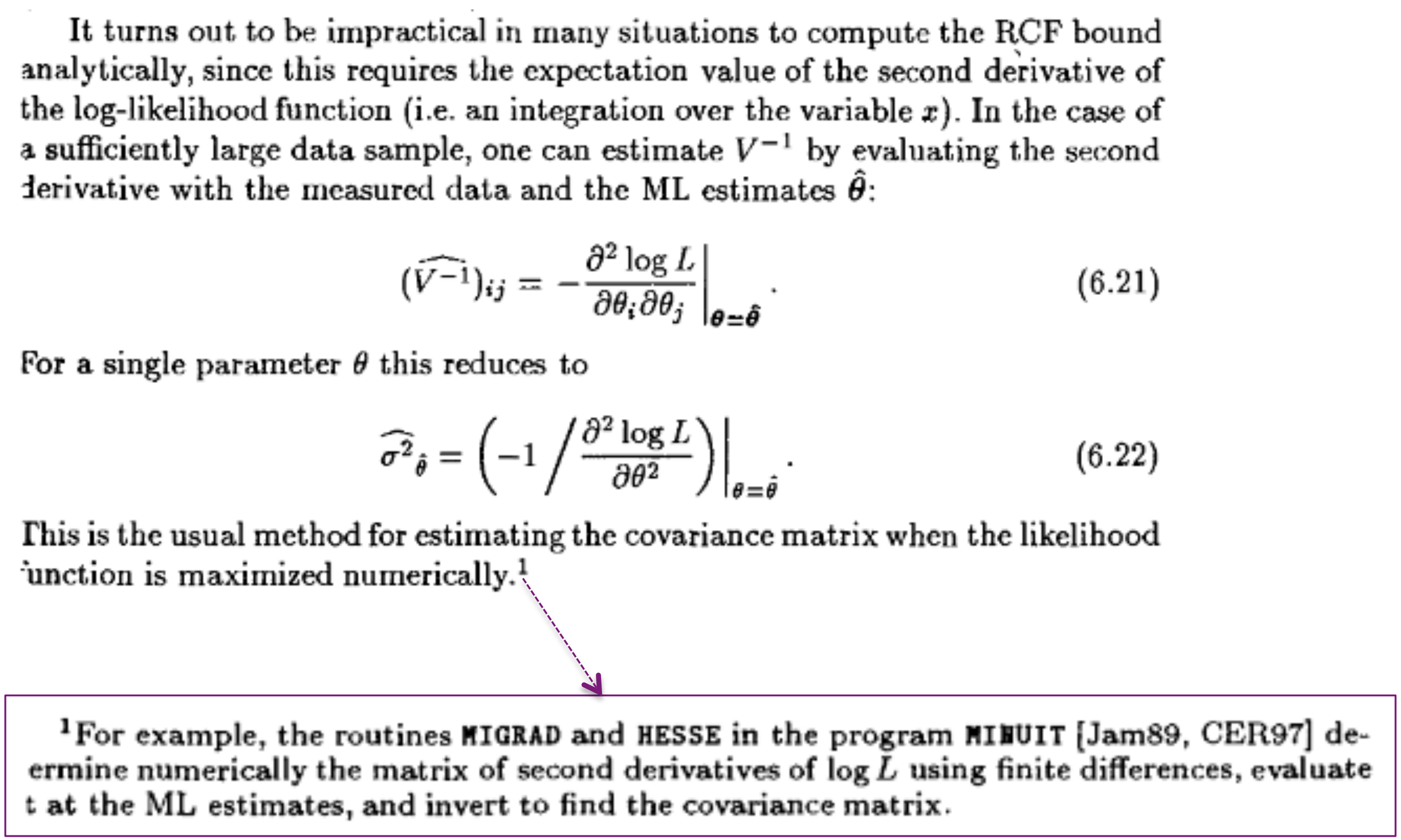
where b is the bias as defined in equation (5.4) and L is the likelihood function.

For the case of more than one parameter, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, the corresponding formula for the inverse of the covariance matrix of their estimators $V_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j]$ is (assuming efficiency and zero bias)

$$(V^{-1})_{ij} = E\left[-\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right] . \qquad (6.19)$$

$$b = E[\hat{\theta}] - \theta .$$

Source: G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998



It turns out to be impractical in many situations to compute the RCF bound analytically, since this requires the expectation value of the second derivative of the log-likelihood function (i.e. an integration over the variable x). In the case of a sufficiently large data sample, one can estimate V^{-1} by evaluating the second derivative with the measured data and the ML estimates $\hat{\boldsymbol{\theta}}$:

$$(\widehat{V^{-1}})_{ij} = - \left. \frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} . \tag{6.21}$$

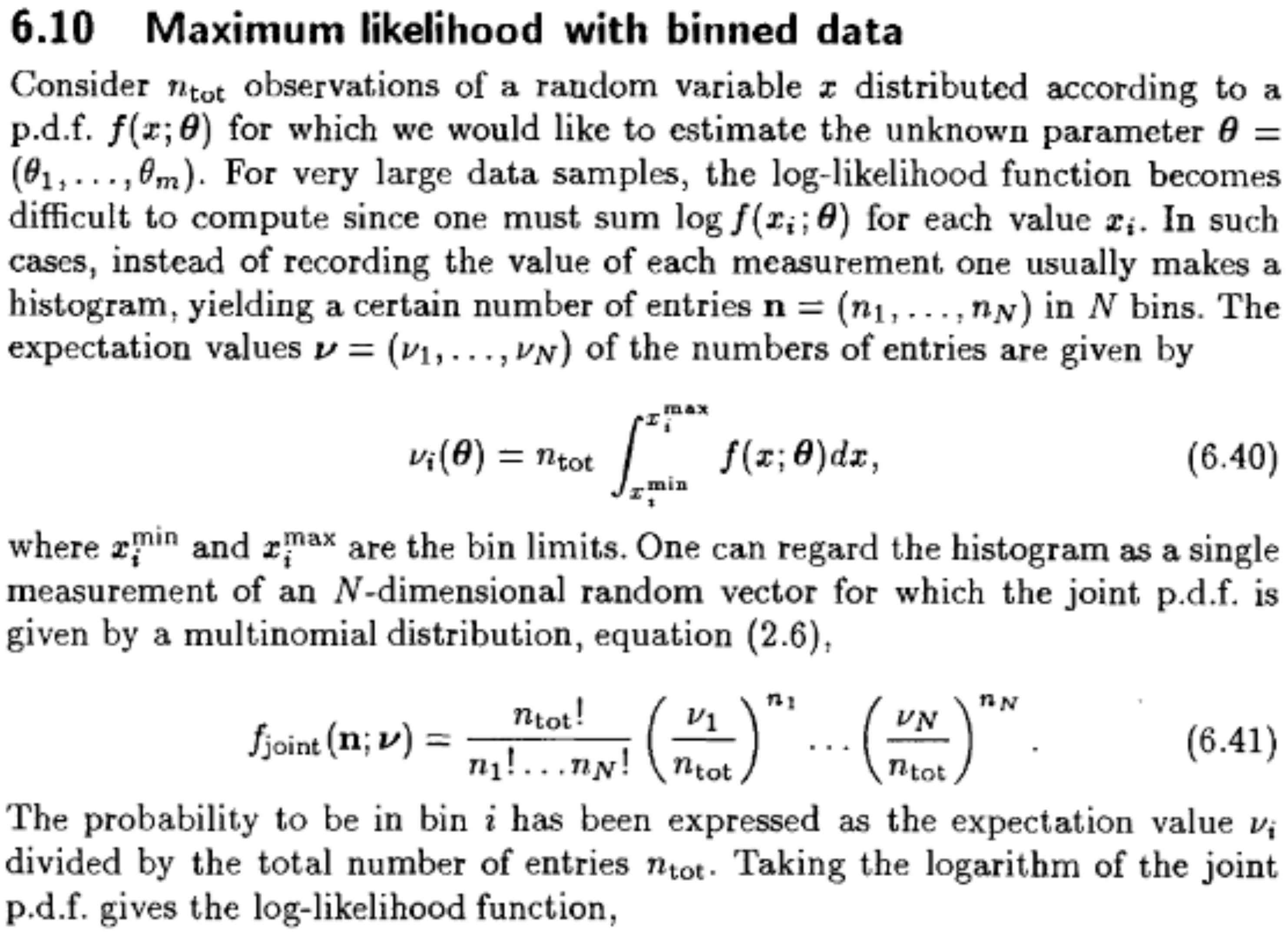
For a single parameter θ this reduces to

$$\widehat{\sigma^2}_{\hat{\theta}} = \left. \left(-1 \left/ \frac{\partial^2 \log L}{\partial\theta^2} \right. \right) \right|_{\theta=\hat{\theta}} . \tag{6.22}$$

This is the usual method for estimating the covariance matrix when the likelihood function is maximized numerically.[1]

[1] For example, the routines MIGRAD and HESSE in the program MINUIT [Jam89, CER97] determine numerically the matrix of second derivatives of $\log L$ using finite differences, evaluate it at the ML estimates, and invert to find the covariance matrix.

Source: G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998



6.10 Maximum likelihood with binned data

Consider n_{tot} observations of a random variable x distributed according to a p.d.f. $f(x;\boldsymbol{\theta})$ for which we would like to estimate the unknown parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. For very large data samples, the log-likelihood function becomes difficult to compute since one must sum $\log f(x_i;\boldsymbol{\theta})$ for each value x_i. In such cases, instead of recording the value of each measurement one usually makes a histogram, yielding a certain number of entries $\mathbf{n} = (n_1, \ldots, n_N)$ in N bins. The expectation values $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$ of the numbers of entries are given by

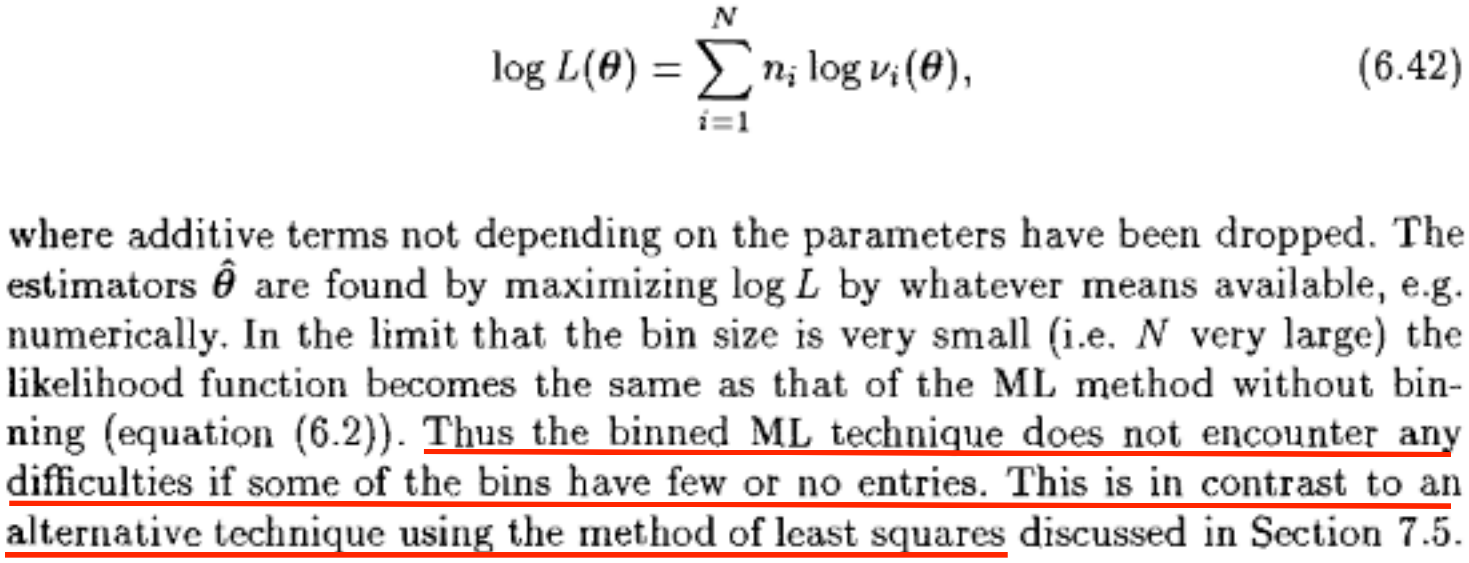
$$\nu_i(\boldsymbol{\theta}) = n_{\text{tot}} \int_{x_i^{\text{min}}}^{x_i^{\text{max}}} f(x;\boldsymbol{\theta})dx, \tag{6.40}$$

where x_i^{min} and x_i^{max} are the bin limits. One can regard the histogram as a single measurement of an N-dimensional random vector for which the joint p.d.f. is given by a multinomial distribution, equation (2.6),

$$f_{\text{joint}}(\mathbf{n};\boldsymbol{\nu}) = \frac{n_{\text{tot}}!}{n_1!\ldots n_N!}\left(\frac{\nu_1}{n_{\text{tot}}}\right)^{n_1} \cdots \left(\frac{\nu_N}{n_{\text{tot}}}\right)^{n_N}. \tag{6.41}$$

The probability to be in bin i has been expressed as the expectation value ν_i divided by the total number of entries n_{tot}. Taking the logarithm of the joint p.d.f. gives the log-likelihood function,

Source: G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998



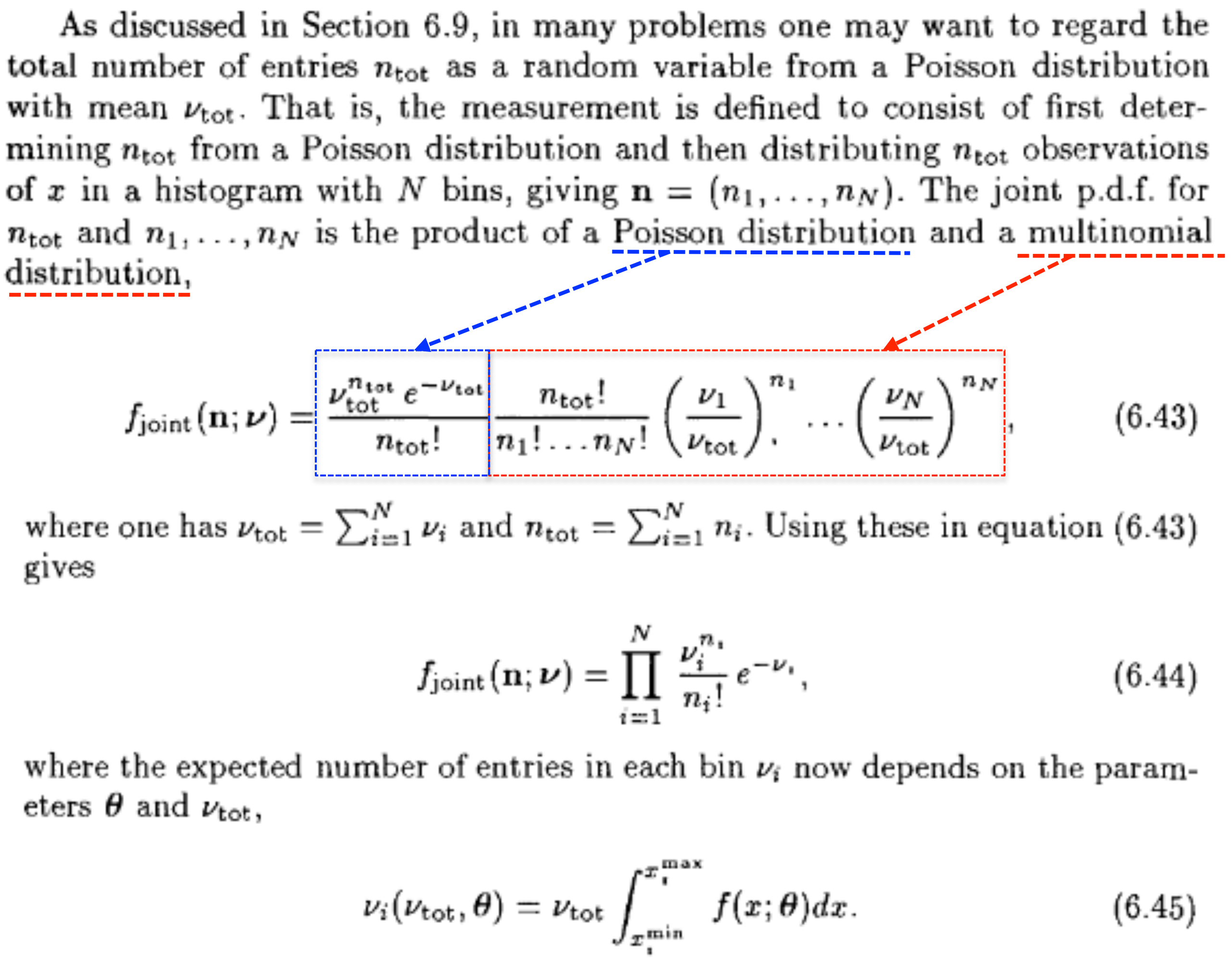
$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} n_i \log \nu_i(\boldsymbol{\theta}), \tag{6.42}$$

where additive terms not depending on the parameters have been dropped. The estimators $\hat{\boldsymbol{\theta}}$ are found by maximizing $\log L$ by whatever means available, e.g. numerically. In the limit that the bin size is very small (i.e. N very large) the likelihood function becomes the same as that of the ML method without binning (equation (6.2)). Thus the binned ML technique does not encounter any difficulties if some of the bins have few or no entries. This is in contrast to an alternative technique using the method of least squares discussed in Section 7.5.

This feature makes Binned ML fit superior than χ^2-fit (LSQ fit)

Source: **G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998**

Extended ML fit in case of binned data



As discussed in Section 6.9, in many problems one may want to regard the total number of entries n_{tot} as a random variable from a Poisson distribution with mean ν_{tot}. That is, the measurement is defined to consist of first determining n_{tot} from a Poisson distribution and then distributing n_{tot} observations of x in a histogram with N bins, giving $\mathbf{n} = (n_1, \ldots, n_N)$. The joint p.d.f. for n_{tot} and $n_1, \ldots, n_N$ is the product of a Poisson distribution and a multinomial distribution,

$$f_{\mathrm{joint}}(\mathbf{n}; \boldsymbol{\nu}) = \boxed{\frac{\nu_{\mathrm{tot}}^{n_{\mathrm{tot}}} e^{-\nu_{\mathrm{tot}}}}{n_{\mathrm{tot}}!}} \boxed{\frac{n_{\mathrm{tot}}!}{n_1! \ldots n_N!} \left(\frac{\nu_1}{\nu_{\mathrm{tot}}}\right)^{n_1} \cdots \left(\frac{\nu_N}{\nu_{\mathrm{tot}}}\right)^{n_N}}, \qquad (6.43)$$

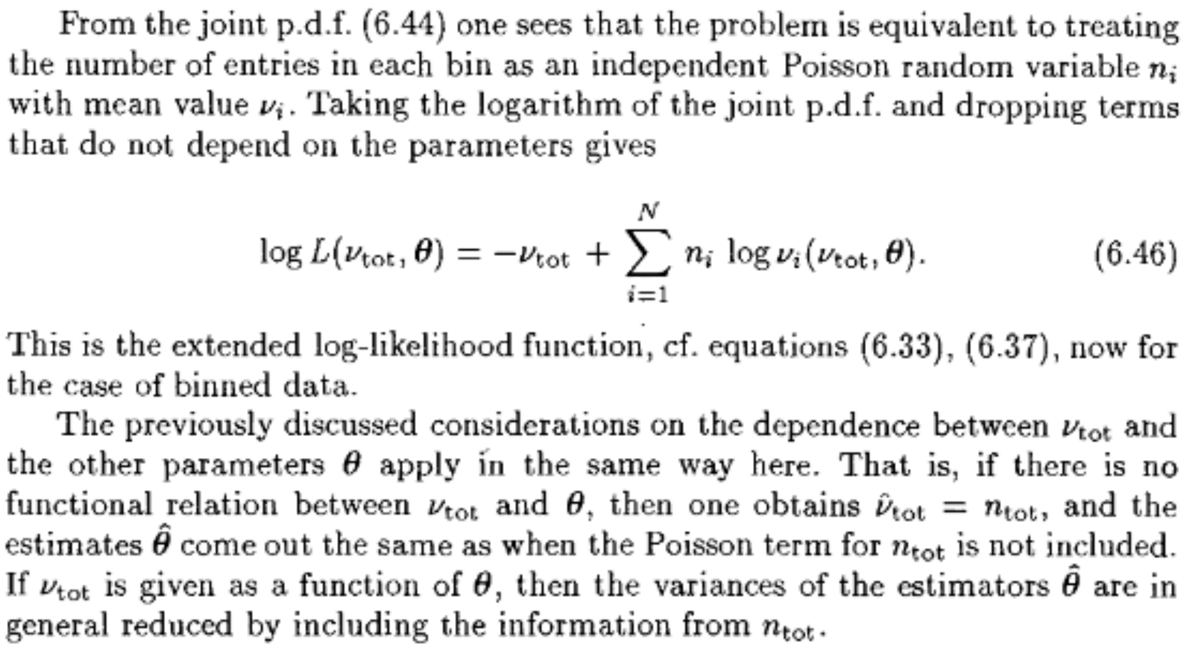
where one has $\nu_{\mathrm{tot}} = \sum_{i=1}^{N} \nu_i$ and $n_{\mathrm{tot}} = \sum_{i=1}^{N} n_i$. Using these in equation (6.43) gives

$$f_{\mathrm{joint}}(\mathbf{n}; \boldsymbol{\nu}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}, \qquad (6.44)$$

where the expected number of entries in each bin ν_i now depends on the parameters $\boldsymbol{\theta}$ and ν_{tot},

$$\nu_i(\nu_{\mathrm{tot}}, \boldsymbol{\theta}) = \nu_{\mathrm{tot}} \int_{x_i^{\mathrm{min}}}^{x_i^{\mathrm{max}}} f(x; \boldsymbol{\theta}) dx. \qquad (6.45)$$

Source: **G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998**



From the joint p.d.f. (6.44) one sees that the problem is equivalent to treating the number of entries in each bin as an independent Poisson random variable n_i with mean value ν_i. Taking the logarithm of the joint p.d.f. and dropping terms that do not depend on the parameters gives

$$\log L(\nu_{\mathrm{tot}}, \boldsymbol{\theta}) = -\nu_{\mathrm{tot}} + \sum_{i=1}^{N} n_i \, \log \nu_i(\nu_{\mathrm{tot}}, \boldsymbol{\theta}). \tag{6.46}$$

This is the extended log-likelihood function, cf. equations (6.33), (6.37), now for the case of binned data.

The previously discussed considerations on the dependence between ν_{tot} and the other parameters $\boldsymbol{\theta}$ apply in the same way here. That is, if there is no functional relation between ν_{tot} and $\boldsymbol{\theta}$, then one obtains $\hat{\nu}_{\mathrm{tot}} = n_{\mathrm{tot}}$, and the estimates $\hat{\boldsymbol{\theta}}$ come out the same as when the Poisson term for n_{tot} is not included. If ν_{tot} is given as a function of $\boldsymbol{\theta}$, then the variances of the estimators $\hat{\boldsymbol{\theta}}$ are in general reduced by including the information from n_{tot}.

Source: **G.Cowan, Statistical Data Analysis, Clarendon Press – Oxford, 1998**