APPENDIX

PROBABILITY THEORY AND STATISTICS

B

CHAPTER OUTLINE

B.1 CRAMÉR-RAO BOUND

Let $\mathbf{x}$ denote a random vector and let $\mathcal{X}$ be a set of corresponding observations, $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$. The corresponding joint pdf is parameterized in terms of the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^l$. The log-likelihood is defined as,

$$L(\boldsymbol{\theta}) := \ln p(\mathcal{X}; \boldsymbol{\theta}).$$

Define the *Fisher's information matrix*

$$J := \begin{bmatrix} \mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_1^2}\right] & \mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2}\right] & \ldots & \mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_l}\right] \\ \vdots & \vdots & \ldots & \vdots \\ \mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_l \partial \theta_1}\right] & \mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_l \partial \theta_2}\right] & \ldots & \mathbb{E}\left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_l^2}\right] \end{bmatrix}. \tag{B.1}$$

Let $I := J^{-1}$ and let $I(i,i)$ denote the ith diagonal element of I. If $\hat{\theta}_i$ is any *unbiased* estimator of the ith component, θ_i, of $\boldsymbol{\theta}$, then the corresponding variance of the estimator,

$$\sigma^2_{\hat{\theta}_i} \geq I(i,i). \tag{B.2}$$

This is known as the Cramér-Rao lower bound, and if an estimator achieves this bound it is said to be *efficient* and it is *unique*.

Machine Learning. http://dx.doi.org/10.1016/B978-0-12-801522-3.09983-3

B.2 CHARACTERISTIC FUNCTIONS

Let $p(x)$ be the probability density function of a random variable x. The associated *characteristic function* is defined as the integral

$$\Phi(\Omega) = \int_{-\infty}^{+\infty} p(x) \exp(j\Omega x)\, dx = \mathbb{E}\left[\exp(j\Omega \mathrm{x})\right]. \tag{B.3}$$

If $j\Omega$ is changed into s, the resulting integral becomes

$$\Phi(s) = \int_{-\infty}^{+\infty} p(x) \exp(sx)\, dx = \mathbb{E}\left[\exp(s\mathrm{x})\right], \tag{B.4}$$

and it is known as the *moment generating function*.

The function

$$\Psi(\Omega) = \ln \Phi(\Omega), \tag{B.5}$$

is known as the *second characteristic function* of x.

The joint characteristic function of l random variables is defined by

$$\Phi(\Omega_1, \Omega_2, \ldots, \Omega_l) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} p(x_1, x_2, \ldots, x_l) \exp\left(j \sum_{i=1}^{l} \Omega_i x_i\right) d\boldsymbol{x}. \tag{B.6}$$

The logarithm of the above is the second joint characteristic function of the l random variables.

B.3 MOMENTS AND CUMULANTS

Taking the nth-order derivative of $\Phi(s)$ in Eq. (B.4) we obtain

$$\frac{d^n \Phi(s)}{ds^n} := \Phi^{(n)}(s) = \mathbb{E}\left[\mathrm{x}^n \exp(s\mathrm{x})\right], \tag{B.7}$$

and hence for $s = 0$

$$\Phi^{(n)}(0) = \mathbb{E}[\mathrm{x}^n] := m_n, \tag{B.8}$$

where m_n is known as the nth-order moment of x. If the moments of all orders are finite, the Taylor series expansion of $\Phi(s)$ near the origin exists and is given by

$$\Phi(s) = \sum_{n=0}^{+\infty} \frac{m_n}{n!} s^n. \tag{B.9}$$

Similarly, the Taylor expansion of the second generating function results in

$$\Psi(s) = \sum_{n=1}^{+\infty} \frac{\kappa_n}{n!} s^n, \tag{B.10}$$

where

$$\kappa_n := \frac{d^n \Psi(0)}{ds^n}, \tag{B.11}$$

and are known as the *cumulants* of the random variable x. It is not difficult to show that $\kappa_0 = 0$. For a zero mean random variable, it turns out that

$$\kappa_1(\mathrm{x}) = \mathbb{E}[\mathrm{x}] = 0, \tag{B.12}$$

$$\kappa_2(\mathrm{x}) = \mathbb{E}[\mathrm{x}^2] = \sigma^2, \tag{B.13}$$

$$\kappa_3(\mathrm{x}) = \mathbb{E}[\mathrm{x}^3], \tag{B.14}$$

$$\kappa_4(\mathrm{x}) = \mathbb{E}[\mathrm{x}^4] - 3\sigma^4. \tag{B.15}$$

That is, the first three cumulants are equal to the corresponding moments. The fourth-order cumulant is also known as *kurtosis*. *For a Gaussian process all cumulants of order higher than two are zero*. The kurtosis is commonly used as a measure of the non-Gaussianity of a random variable. For random variables described by (unimodal) pdfs with spiky shapes and heavy tails, known as leptokurtic or super-Gaussian, κ_4 is positive, whereas, for random variables associated with pdfs with a flatter shape, known as platykurtic or sub-Gaussian, κ_4 is negative. Gaussian variables have zero kurtosis. The opposite is not always true, in the sense that there exist non-Gaussian random variables with zero kurtosis; however, this can be considered rare.

Similar arguments hold for the expansion of the joint characteristic functions for multivariate pdfs. For zero mean random variables, $\mathrm{x}_i, i = 1, 2, \ldots, l$, the cumulants of order up to four are given by

$$\kappa_1(\mathrm{x}_i) = \mathbb{E}[\mathrm{x}_i] = 0, \tag{B.16}$$

$$\kappa_2(\mathrm{x}_i, \mathrm{x}_j) = \mathbb{E}[\mathrm{x}_i\mathrm{x}_j], \tag{B.17}$$

$$\kappa_3(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k) = \mathbb{E}[\mathrm{x}_i\mathrm{x}_j\mathrm{x}_k], \tag{B.18}$$

$$\kappa_4(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k, \mathrm{x}_r) = \mathbb{E}[\mathrm{x}_i\mathrm{x}_j\mathrm{x}_k\mathrm{x}_r] - \mathbb{E}[\mathrm{x}_i\mathrm{x}_j]\,\mathbb{E}[\mathrm{x}_k\mathrm{x}_r] \tag{B.19}$$

$$- \mathbb{E}[\mathrm{x}_i\mathrm{x}_k]\,\mathbb{E}[\mathrm{x}_j\mathrm{x}_r] - \mathbb{E}[\mathrm{x}_i\mathrm{x}_r]\,\mathbb{E}[\mathrm{x}_j\mathrm{x}_k]. \tag{B.20}$$

Thus, once more, the cumulants of the first three orders are equal to the corresponding moments. If all variables coincide, we talk about *auto-cumulants*, and otherwise about *cross-cumulants*, i.e.,

$$\kappa_4(\mathrm{x}_i, \mathrm{x}_i, \mathrm{x}_i, \mathrm{x}_i) = \kappa_4(\mathrm{x}_i),$$

that is, the fourth order auto-cumulant of x_i is identical to its kurtosis. It is not difficult to see that if the zero mean random variables are mutually independent, their cross-cumulants are zero. *This is also true for the cross-cumulants of all orders*.

B.4 EDGEWORTH EXPANSION OF A PDF

Taking into account the expansion in Eq. (B.10), the definition given in Eq. (B.5), and taking the inverse Fourier of $\Phi(\Omega)$ in Eq. (B.3) we can obtain the following expansion of $p(x)$ for a zero mean unit variance random variable x:

$$p(x) = g(x)\bigg(1 + \frac{1}{3!}\kappa_3(\mathrm{x})H_3(x) + \frac{1}{4!}\kappa_4(\mathrm{x})H_4(x) + \frac{10}{6!}\kappa_3^2(\mathrm{x})H_6(x)$$

$$+ \frac{1}{5!}\kappa_5(\mathrm{x})H_5(x) + \frac{35}{7!}\kappa_3(\mathrm{x})\kappa_4(\mathrm{x})H_7(x) + \ldots\bigg), \tag{B.21}$$

where $g(x)$ is the unit variance and zero mean normal pdf, and $H_\kappa(x)$ is the Hermite polynomial of degree k. The rather strange ordering of terms is the outcome of a specific reordering in the resulting expansion, so that the successive coefficients in the series decrease uniformly. This is very important when truncation of the series is required. The Hermite polynomials are defined as

$$H_k(x) = (-1)^k \exp(x^2/2) \frac{d^k}{dx^k} \exp(-x^2/2), \tag{B.22}$$

and they form a complete orthogonal basis set in the real axis, i.e.,

$$\int_{-\infty}^{+\infty} \exp(-x^2/2) H_n(x) H_m(x)\, dx = \begin{cases} n!\sqrt{2\pi} & \text{if } n = m \\ 0 & \text{if } n \neq m. \end{cases} \tag{B.23}$$

The expansion of $p(x)$ in Eq. (B.21) is known as the *Edgeworth expansion*, and it is actually an expansion of a pdf around the normal one [1].

REFERENCE

[1] A. Papoulis, A.U. Pillai, Probability, Random Variables and Stochastic Processes, fourth ed., McGaw Hill, New York, 2002.