

#Principles and pathology of orthodox statistics.

← Back to Chapters

Comments on 17.1.

Given a fixed prior and model, a Bayesian estimator (MAP, posterior mean, or minimizing some other posterior risk instead of MSE, or of any type at all) is still an estimator - it is a number $f(X)$ produced from data X . It still partitions the data manifold into parts on which it gives the same estimate. Thus this line of Jaynes's reasoning does not seem to distinguish between frequentist ("orthodox") and Bayesian procedures.

Formula 17.25

"This decreases monotonically from 1 at $n = 0$ to $1/2$ as $n \rightarrow \infty$ " – see this [MathOverflow](#) answer for a confirmation of monotonicity.

Comments on 17.4

Short summary of derivation of Cramer-Rao bound 17.30: apply Cauchy-Schwartz to the "covariance dot-product" of the score and the estimator. Slightly longer summary is as follows.

The formula 17.29 just expresses the following: covariance is a dot product (positive definite bilinear form) on the space of zero-mean random variables with finite second moments; the "(Cauchy-Buniakowski-)Schwartz" inequality for this dot product aka $u \cdot v \leq |u||v|$ is the formula 17.29. Attempting to embed random variables into L^2 functional space lead to the slightly-awkward formulas 17.28.

The cool thing is that the score – the derivative (in the parameter) of the log-likelihood is – a zero mean random variable, whose variance is the Fisher information, and whose covariance with an estimator β is the derivative (in the parameter) of the estimator's expectation. This is why we get that standard deviation of the estimator is bounded below by (1+derivative of bias) over root of the Fisher information.

The equality condition is that the estimator and the score are proportional: that is, that the parameter is the parameter of exponential family, aka the "maximal entropy" family.

Comments on 17.5.1

Here is how Jaynes's criticism of "rejection of null hypothesis" statistical method transfers to a criticism of mathematical technique of "proof by contradiction"

(quote from 15.5.1 with italics for modifications)"

This is the kind of logic that underlies *all proofs by contradiction*. In order to argue for a *conclusion* C , one does it indirectly: invent a hypothesis "*not* C " that denies C , then argue against "*not* C " in a way that makes no reference to C at all (that is, using only *deductions based on* "*not* C "). To see how far this procedure takes us from elementary logic, suppose we decide that the *conclusion holds*; that is, we reject "*not* C ". Surely, we must also reject *reasoning* based on "*not* C "; but then what was the logical justification for *the deduction*? *Proof by contradiction* saws off its own limb.