The 7th international conference on Geometric Science of Information

Honorable Mention

Category: Questioning History

PRESENTED TO

Olivier Rioul

AUTHOR OF THE PAPER

A Historical Perspective on the Schützenberger-van-Trees Inequality: A Posterior Uncertainty Principle

GSI 2025
October 29th-31st 2025
Saint-Malo

Frédéric Barbaresco
Thales, France
General co-chair

Frank Nielsen
Sony Computer Science Laboratories, Japan
General co-chair

Alice Barbora Tumpach
CNRS Pauli Institute, Austria
Lille University, France
Prize committee, President

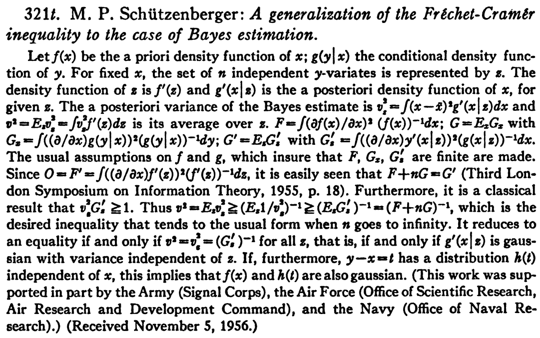
HONORABLE MENTION in the category QUESTIONING HISTORY

Olivier Rioul

Olivier Rioul (https://perso.telecom-paristech.fr/rioul/) is Full Professor at the Department of Communication and Electronics at Télécom Paris, Institut Polytechnique de Paris, France. He graduated from École Polytechnique and from École Nationale Supérieure des Télécommunications, Paris, France, where he obtained his PhD degree. His research interests are in applied mathematics and include various, sometimes unconventional, applications of information theory such as inequalities in statistics, hardware security, and experimental psychology. He has been teaching information theory and statistics at various universities for twenty years and has published a textbook which has become a classical French reference in the field.

A Historical Perspective on the Schützenberger-van-Trees Inequality: A Posterior Uncertainty Principle

The Bayesian Cramér-Rao Bound (BCRB) is generally attributed to Van Trees who published it in 1968. According to Stigler's law of eponymy, no scientific discovery is named after its first discoverer. This is the case not only for the Cramér-Rao bound itself—due in particular to the French mathematicians Fréchet and Darmois—but also for the van Trees inequality: The French physician, geneticist, epidemiologist and mathematician Marcel-Paul (Marco) Schützenberger, in a paper of just fifteen lines written in 1956 (see picture)—more than a decade before van Trees—had not only derived the BCRB but, as a close examination of his proof shows, used a very original approach based on the Weyl-Heisenberg uncertainty principle on the square root of the posterior distribution. The work of Olivier Rioul is not only correcting the historical facts surrounding the Schützenberger-Van Trees inequality but also drawing inspiration from the Schützenberger's original idea to propose new developments and extends Schützenberger's approach to Fisher information matrices, which opens up new perspectives.



321*t*. M. P. Schützenberger: *A generalization of the Fréchet-Cramér inequality to the case of Bayes estimation.*

Let $f(x)$ be the a priori density function of x; $g(y|x)$ the conditional density function of y. For fixed x, the set of n independent y-variates is represented by z. The density function of z is $f'(z)$ and $g'(x|z)$ is the a posteriori density function of x, for given z. The a posteriori variance of the Bayes estimate is $v_z^2=\int(x-\bar{x})^2g'(x|z)dx$ and $v^2=E_zv_z^2=\int v_z^2f'(z)dz$ is its average over z. $F=\int(\partial f(x)/\partial x)^2\,(f(x))^{-1}dx$; $G=E_xG_x$ with $G_x=\int((\partial/\partial x)g(y|x))^2(g(y|x))^{-1}dy$; $G'=E_zG'_z$ with $G'_z=\int((\partial/\partial x)y'(x|z))^2(g(x|z))^{-1}dx$. The usual assumptions on f and g, which insure that F, G_x, G'_z are finite are made. Since $O=F'=\int((\partial/\partial x)f'(z))^2(f'(z))^{-1}dz$, it is easily seen that $F+nG=G'$ (Third London Symposium on Information Theory, 1955, p. 18). Furthermore, it is a classical result that $v_z^2G'_z\geqq 1$. Thus $v^2=E_zv_z^2\geqq(E_z1/v_z^2)^{-1}\geqq(E_zG'_z)^{-1}=(F+nG)^{-1}$, which is the desired inequality that tends to the usual form when n goes to infinity. It reduces to an equality if and only if $v^2=v_z^2=(G'_z)^{-1}$ for all z, that is, if and only if $g'(x|z)$ is gaussian with variance independent of z. If, furthermore, $y-x=t$ has a distribution $h(t)$ independent of x, this implies that $f(x)$ and $h(t)$ are also gaussian. (This work was supported in part by the Army (Signal Corps), the Air Force (Office of Scientific Research, Air Research and Development Command), and the Navy (Office of Naval Research).) (Received November 5, 1956.)

Theorem 1 (Schützenberger's Inequality (BCRB)). *Let $X|\Theta$ be a regular Bayesian statistical model. The quadratic (mean-squared error) risk* $\mathbf{R}\triangleq\mathbb{E}_{x,\theta}\big\{(\hat\theta(X)-\theta)(\hat\theta(X)-\theta)^t\big\}$ *is lower bounded (in Loewner order's sense) by the inverse of the joint Fisher information matrix* $\mathbf{J}\triangleq\mathbb{E}_{x,\theta}\big\{\nabla\log p(X,\theta)\nabla^t\log p(X,\theta)\big\}$:

$$\boxed{\mathbf{R}\geq\mathbf{J}^{-1}}$$

Proof. It is well known that the quadratic (mean-squared error) risk is minimized for the MMSE estimator, given by the mean of the posterior distribuition $\hat\theta^*(x)=\mathbb{E}(\theta|x)$. Therefore, it suffices to prove the inequality on the minimal risk $\min\mathbf{R}=\mathbb{E}_x\operatorname{Cov}(\theta|x)$, where $\operatorname{Cov}(\theta|x)=\mathbb{E}_{\theta|x}\{(\theta-\mathbb{E}(\theta|x))(\theta-\mathbb{E}(\theta|x))^t\}$ is the covariance matrix of the *posterior*. The (matrix) Weyl-Heisenberg inequality (a.k.a. uncertainty principle) $\mathbf{R}_{t\cdot f}\geq\frac14\mathbf{R}^{-1}_{\nabla f}$ applied to the function $f(\theta)=\sqrt{p(\theta|x)}$ for fixed x, reads, after making a change of variable $\theta\leftarrow\theta-\mathbb{E}(\theta|x)$, $\operatorname{Cov}(\theta|x)\geq\frac14\mathbf{R}^{-1}_{\nabla\sqrt{p(\theta|x)}}$. Now since $\nabla\sqrt{p(\theta|x)}=\frac{1}{2\sqrt{p(\theta|x)}}\nabla p(\theta|x)$, we have $\mathbf{R}_{\nabla\sqrt{p(\theta|x)}}=\frac14\mathbb{E}_{\theta|x}\big\{\nabla\log p(\theta|x)\nabla^t\log p(\theta|x)\big\}=\frac14\tilde{\mathbf{J}}(x)$, which gives $\operatorname{Cov}(\theta|x)\geq\tilde{\mathbf{J}}(x)^{-1}$ for any fixed data vector x, where the *posterior Fisher information matrix*: $\tilde{\mathbf{J}}(x)\triangleq\mathbb{E}_{\theta|x}\big\{\nabla\log p(\theta|x)\nabla^t\log p(\theta|x)\big\}$ satisfies the relation $\mathbf{J}=\mathbb{E}_x\tilde{\mathbf{J}}(x)$, as is easily checked. Taking the expectation over the unconditional law $p(x)$ and applying the operator convexity of the function $A\mapsto A^{-1}$ concludes: $\mathbf{R}\geq\mathbb{E}_x\operatorname{Cov}(\theta|x)\geq\mathbb{E}_x\big(\tilde{\mathbf{J}}(x)^{-1}\big)\geq\big(\mathbb{E}_x\tilde{\mathbf{J}}(x)\big)^{-1}=\mathbf{J}^{-1}$.