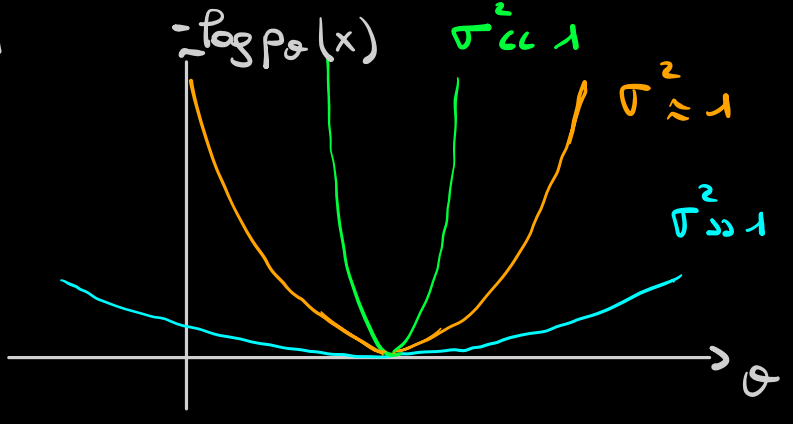
Fisher information and Cramér-Rao lower bound

Example (Gaussians with given variance). $X_j \overset{iid}{\sim} \mathcal{N}(\vartheta, \sigma^2)$

$\Theta = \mathbb{R}$, σ^2 is known

Consider the estimator $\hat{\vartheta} = \frac{1}{n}\sum X_j$ $\longrightarrow$ UMVU!

$$\mathbb{V}_\vartheta(\hat{\vartheta}) = \frac{1}{n}\cdot\sigma^2 = \frac{1}{n}\left(\frac{1}{\sigma^2}\right)^{-1}$$



Idea: the curvature of the (log) likelihood around the estimator

$$-\frac{\partial^2}{\partial\vartheta^2}\log p_\vartheta(x) = \frac{1}{\sigma^2}$$

indicates how "localized" the parameter ϑ is. ($\Rightarrow$ how much variance/uncertainty our estimate might have).

Averaging over $x \sim \mathbb{P}_\vartheta$ we obtain

$$I(\vartheta) = -\mathbb{E}_\vartheta\left(\frac{\partial^2}{\partial\vartheta^2}\log p_\vartheta(x)\right)$$

We expect (the inverse of) this quantity to encode how much variance we may expect our estimator (at least).

Idea: We want to capture the variability of a parametric model around ϑ.

Assumption 1: Let $(\mathcal{X}, \mathcal{F}, \{P_\vartheta\}_{\vartheta\in\Theta})$ be a dominated statistical model with densities $p_\vartheta(x)$ wrt. a dominating meas. μ. We assume that $\Theta \subseteq \mathbb{R}^q$ is open and that

$$p_\vartheta(x) > 0 \qquad \forall x \in \mathcal{X} \quad \forall \vartheta \in \Theta$$

$\mathbb{E}$ denotes expectation wrt μ.

→ to take logs

Ex: Gaussians: $\vartheta = (\mu, \sigma^2)$ $X_j \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\Rightarrow p_\vartheta(x) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_j-\mu)^2}{2\sigma^2}\right) > 0 \qquad \forall x \in \mathbb{R}^n$$

Assumption 2: It holds that for all mble f with $\mathbb{E}(f^2(x)) < \infty$

$$\nabla_\vartheta \mathbb{E}\left(p_\vartheta(x) f(x)\right) = \mathbb{E}\left(\nabla_\vartheta p_\vartheta(x) f(x)\right)$$

(for example: $\nabla_\vartheta p_\vartheta(x)$ exists and $\mathbb{E}(\|\nabla_\vartheta p_\vartheta(x)\|^2) < \infty$ $\forall \vartheta \in \Theta$)

Lemma: Under A1, A2 we have

$$\mathbb{E}_\vartheta\left(\nabla_\vartheta \log p_\vartheta(x)\right) = 0 \qquad \forall \vartheta \in \Theta$$

Proof: $\mathbb{E}_\vartheta\left(\nabla_\vartheta \log p_\vartheta(x)\right) = \mathbb{E}_\vartheta\left(\frac{1}{p_\vartheta(x)} \cdot \nabla_\vartheta p_\vartheta(x)\right) =$

$$= \int \frac{1}{p_\vartheta(x)} \nabla_\vartheta p_\vartheta(x) \cdot p_\vartheta(x)\, \mu(dx)$$

$$= \int \nabla_\vartheta p_\vartheta(x)\, \mu(dx) = \nabla_\vartheta \int p_\vartheta(x)\, \mu(dx)$$

$$= \nabla_\vartheta(1) = 0 \,. \qquad \square$$

Assumption 3: $\mathbb{E}_\vartheta\left(\|\nabla_\vartheta \log p_\vartheta(x)\|^2\right) < \infty \quad \forall \vartheta \in \Theta$

Def: Let A1,2,3 hold. We call

$$I(\vartheta)_{ij} = \mathbb{E}_\vartheta\left(\frac{\partial}{\partial \vartheta_i}\log p_\vartheta(x) \cdot \frac{\partial}{\partial \vartheta_j}\log p_\vartheta(x)\right) \qquad i,j \in \{1,\dots k\}$$

the Fisher infomation matrix of the dominated model. $\{p_\vartheta(x)\}_{\vartheta\in\Theta}$. Further, we call $s_\vartheta(x) = \nabla_\vartheta \log p_\vartheta(x)$ the score function.

Note: Alternatively, using chain rule, we can write

$$I(\vartheta)_{ij} = \mathbb{E}_\vartheta\left(\frac{1}{p_\vartheta(x)^2}\frac{\partial}{\partial \vartheta_i}p_\vartheta(x)\frac{\partial}{\partial \vartheta_j}p_\vartheta(x)\right)$$

Note: The matrix $I(\vartheta)_{ij}$ is symmetric positive semidefinite since it is the covariance matrix C of the score:

$$\begin{aligned} \langle C\cdot v, v\rangle &= \sum_{i,j=1}^{k} \mathrm{Cov}(Y_i, Y_j)\, v_i v_j \\ &= \mathbb{E}\left(\sum_{i,j=1}^{k} v_i (Y_i - \mathbb{E}Y_i)(Y_j - \mathbb{E}(Y_j))v_j\right) \\ &= \mathbb{E}\left(\left(\sum v_i (Y_i - \mathbb{E}(Y_i))\right)^2\right) \geq 0 \end{aligned}$$

under A1-3

Lemma: For every $i,j \in \{1,\dots k\}$ $\quad \mathbb{E}_\vartheta\left(\frac{1}{p_\vartheta(x)}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}p_\vartheta(x)\right) = 0$

Proof: $\mathbb{E}_\vartheta\left(\frac{1}{p_\vartheta(x)}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}p_\vartheta(x)\right) = \int \frac{1}{p_\vartheta(x)}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}p_\vartheta(x)\, p_\vartheta(x)\,\mu(dx)$

$$= \frac{\partial^2}{\partial\vartheta_i \partial\vartheta_j} \int p_\vartheta(x)\mu(dx) = \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j} 1 = 0$$ □

Proposition: Under A1-3

$$I(\vartheta)_{ij} = -\mathbb{E}_\vartheta\left(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j} \log p_\vartheta(x)\right).$$

Proof: Use the notation $\partial_{ij}^2 = \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}$.

$$\partial_{ij}^2 \log p_\vartheta(x) = \partial_i\left(\partial_j \log p_\vartheta(x)\right) = \partial_i\left(\frac{1}{p_\vartheta(x)}\partial_j p_\vartheta(x)\right)$$

$$= \frac{1}{p_\vartheta(x)}\partial_{ij}^2 p_\vartheta(x) - \frac{1}{p_\vartheta(x)^2}\partial_i p_\vartheta(x)\,\partial_j p_\vartheta(x)$$

$$\Rightarrow \mathbb{E}_\vartheta\left(\partial_{ij}\log p_\vartheta(x)\right) = \mathbb{E}_\vartheta\left(\frac{1}{p_\vartheta(x)}\partial_{ij}^2 p_\vartheta(x)\right) - \mathbb{E}_\vartheta\left(\frac{1}{p_\vartheta(x)}\partial_i p_\vartheta(x)\partial_j p_\vartheta(x)\right)$$

$$= -I(\vartheta)_{ij}$$ □

Ex (Gaussian) $\underline{X} = (X_1,\dots X_n)$ $X_j \overset{iid}{\sim} \mathcal{N}(m,\sigma^2)$ $\vartheta = \mu$.

$$\log p_\vartheta(x) = \sum_{j=1}^n -\frac{1}{2}\log 2\pi\sigma^2 - (x_j-m)^2\frac{1}{2\sigma^2}$$

$$\Rightarrow \partial_\vartheta \log p_\vartheta(x) = +\sum_{j=1}^n \frac{x_j-m}{\sigma^2} \Rightarrow -\partial_\vartheta^2 \log p_\vartheta(x) = n\cdot\frac{1}{\sigma^2}$$

$$\Rightarrow \mathbb{E}_\vartheta\left(\left(\partial_\vartheta \log p_\vartheta(x)\right)^2\right) = \frac{1}{\sigma^4}\sum_{j=1}^n \mathbb{E}_\vartheta\left((x_j-m)^2\right) = n\frac{1}{\sigma^2}$$

Now consider the statistical model $(\mathbb{X}^n, \mathcal{F}^{\otimes n}, \{\mathbb{P}_\vartheta^{\otimes n}\}_{\vartheta\in\Theta})$.

Theorem: Let A1-3 hold and consider the $\mu_1\otimes\mu_2$ dominated stat. model $(\mathbb{X}_1\times\mathbb{X}_2, \mathcal{F}_1\otimes\mathcal{F}_2, (\mathbb{P}_\vartheta^{(1)}\otimes\mathbb{P}_\vartheta^{(2)})_{\vartheta\in\Theta})$ then

$$I^{(1+2)}(\vartheta) = I^{(1)}(\vartheta) + I^{(2)}(\vartheta)$$

where $I_c(\vartheta)_{ij} = \mathbb{E}_\vartheta^{(c)}\left(\frac{\partial}{\partial\vartheta_i}\log p_\vartheta^{(c)}(x)\,\frac{\partial}{\partial\vartheta_j}\log p_\vartheta^{(c)}(x)\right)$.

Corollary: for $(\mathbb{X}^n, \mathcal{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n})_\vartheta)$ $\quad I_n(\vartheta) = n\, I_1(\vartheta)$

↳ fisher of $\mathbb{P}_\vartheta$

⟹ compute only $I_1(\vartheta)$ ↳ fisher of $\mathbb{P}_\vartheta^{\otimes n}$

Ex (Gaussian unknown mean) $X_j \overset{iid}{\sim} \mathcal{N}(m,\sigma^2)$ $\quad \vartheta = m$.

$$I_1(\vartheta) = -\mathbb{E}_\vartheta\left(\partial_\vartheta^2 \log p_\vartheta(x)\right) = \frac{1}{\sigma^2} \longrightarrow I_n(\vartheta) = n\cdot\frac{1}{\sigma^2}$$

Ex (Gaussian unknown variance) $X_j \overset{iid}{\sim} \mathcal{N}(m,\sigma^2)$ $\quad \vartheta = \sigma^2$

$$\log p_\vartheta(x) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{(x-m)^2}{2\sigma^2} \longrightarrow s_\vartheta(x) = -\frac{1}{2\sigma^2} + \frac{(x-m)^2}{2\sigma^4}$$

$$I_1(\vartheta) = -\mathbb{E}_\vartheta\left(\partial_\vartheta^2 \log p_\vartheta(x)\right) = -\mathbb{E}_\vartheta\left(\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-m)^2\right)$$

$$= -\left(\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\sigma^2\right) = \frac{1}{2\sigma^4} \Longrightarrow I_n(\vartheta) = \frac{n}{2\sigma^4}$$

Exercise: $X_j \sim \mathcal{N}(m,\sigma^2)$, $\vartheta = (m,\sigma^2)$ $\quad I(\vartheta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$

Lemma: Let $X \perp\!\!\!\perp Y$ random vectors, then $\mathrm{Cov}(X+Y) = \mathrm{Cov}(X) + \mathrm{Cov}(Y)$

Proof of theorem:

$$I(\vartheta) = \mathrm{Cov}\left(\nabla_\vartheta \log p_\vartheta(x)\right) = \mathrm{Cov}\left(\nabla_\vartheta\left(\log p_\vartheta^{(1)}(x) + \log p_\vartheta^{(2)}(x)\right)\right)$$

$$\overset{\text{Lemma}}{=} \mathrm{Cov}\left(\nabla \log p_\vartheta^{(1)}(x)\right) + \mathrm{Cov}\left(\nabla \log p_\vartheta^{(2)}(x)\right) = I_1(x) + I_2(x) \qquad \square$$

Theorem (Cramer-Rao) Let T be an unbiased estimator of $g(\vartheta)$ with $\mathbb{E}_\vartheta(T^2) < \infty \ \forall \vartheta \in \Theta$. Then, assuming I is invertible $\forall \vartheta$,

$$V_\vartheta(T) \geq \nabla_\vartheta g(\vartheta)^T \left(I(\vartheta)\right)^{-1} \nabla_\vartheta g(\vartheta)$$

Def: We call $\nabla_\vartheta g(\vartheta)^T \left(I(\vartheta)\right)^{-1} \nabla_\vartheta g(\vartheta)$ the Cramer-Rao lower bound for estimating $g(\vartheta)$

Note: in 1d $\quad V_\vartheta(T) \geq \dfrac{\left(\partial_\vartheta g(\vartheta)\right)^2}{I(\vartheta)}$

if $g(\vartheta) = \vartheta \quad V_\vartheta(T) \geq I(\vartheta)^{-1}$

Proof: We have

$$\nabla_\vartheta g(\vartheta) = \nabla_\vartheta \mathbb{E}_\vartheta(T) = \nabla_\vartheta \int T(x)\, p_\vartheta(x)\, \mu(dx)$$

$$= \int T(x)\, \nabla_\vartheta p_\vartheta(x)\, \frac{1}{p_\vartheta(x)}\, p_\vartheta(x)\, \mu(dx)$$

$$= \mathbb{E}_\vartheta\left(T(x)\, \nabla_\vartheta \log p_\vartheta(x)\right) - \mathbb{E}_\vartheta(T(x))\, \mathbb{E}_\vartheta\left(\nabla_\vartheta \log p_\vartheta(x)\right)$$

$$= \mathbb{E}_\vartheta\left(\nabla_\vartheta \log p_\vartheta(x)\, \left(T(x) - \mathbb{E}_\vartheta(T(x))\right)\right).$$

$$\Longrightarrow \quad \langle x, \nabla_\vartheta E_\vartheta(T)\rangle^2 = E_\vartheta\left(\langle x, \nabla_\vartheta \log p_\vartheta(x)\rangle (T(x) - E_\vartheta(T(x)))\right)$$

$$\overset{C.S}{\leq} V_\vartheta(T)\, E\left(\langle x, \nabla_\vartheta \log p_\vartheta(x)\rangle^2\right) =$$

$$= V_\vartheta(T)\; x^T I(\vartheta) x .$$

$\longrightarrow$ done in the 1d case: simplify x and get $V(T) \geq \frac{(\partial_\vartheta g(\vartheta))^2}{I(\vartheta)}$.

$$\Longrightarrow \quad V_\vartheta(T) \geq \frac{\langle x, \nabla_\vartheta E_\vartheta(T)\rangle^2}{x^T I(\vartheta) x}$$

$$\overset{\text{optimize over } x}{\rightsquigarrow} \quad V_\vartheta(T) = \nabla_\vartheta g(\vartheta)^T I(\vartheta)^{-1} \nabla_\vartheta g(\vartheta)$$

Lemma: Let $a \in \mathbb{R}^q$, A symmetric positive definite, then

$$\sup_{x \in \mathbb{R}^q} \frac{\langle x, a\rangle^2}{\langle Ax, x\rangle} = \langle A^{-1} a, a\rangle$$

Pf: Let $BB = A$, B symmetric $\longrightarrow$

$$\frac{\langle x, a\rangle}{\langle Ax, x\rangle} = \frac{\langle B^{-1} y, a\rangle}{\langle AB^{-1}y, B^{-1}y\rangle} = \frac{\langle y, B^{-1}a\rangle}{\|y\|^2} = \left(\frac{\langle y, b\rangle}{\|y\|}\right)^2 = \|b\|^2 = \langle (B^2)^{-1} a, a\rangle$$

$$= \langle A^{-1} a, a\rangle \qquad \square$$

$\longrightarrow$ maximized in $y = b$

Ex (Poisson): $\underline{X} = (X_1, \dots X_n)$ $\quad X_j \overset{iid}{\sim} \text{Pois}(\vartheta)$ $\quad \vartheta > 0$

$$\log p_\vartheta(x) = -\vartheta + x \log \vartheta - \log x!$$

$$\Rightarrow s_\vartheta(x) = \nabla_\vartheta \log p_\vartheta(x) = -1 + \frac{x}{\vartheta}$$

$$\Rightarrow I_1(\vartheta) = V_\vartheta\left(s_\vartheta(x)\right) = V_\vartheta\left(\frac{x}{\vartheta}\right) = \frac{1}{\vartheta^2} V_\vartheta(x) = \frac{1}{\vartheta}.$$

For $\hat{\vartheta} = \frac{1}{n}\sum x_j$, $g(\vartheta) = \vartheta$

$$V_\vartheta(\hat{\vartheta}) = \frac{1}{n}\vartheta = \frac{1}{n \cdot \frac{1}{\vartheta}} = \frac{(\partial_\vartheta g(\vartheta))^2}{I_n(\vartheta)}$$

CRLB.

Now for $g(\vartheta) = e^{-\vartheta}$, $T = \left(\frac{n-1}{n}\right)^{\sum x_j}$ (UMVU)

$$E_\vartheta(T^2) = \sum_{j=0}^{\infty} \left(1 - \frac{1}{n}\right)^{2j} \frac{(n\vartheta)^j}{j!} e^{-n\vartheta}$$

$$= e^{-n\vartheta} \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{(n-1)^2}{n}\vartheta\right)^j = e^{-n\vartheta} e^{\frac{(n-1)^2\vartheta}{n}}$$

$$= \exp\left(\frac{(1-2n)\vartheta}{n}\right)$$

$$\Rightarrow V_\vartheta(T) = E_\vartheta T^2 - \left(E_\vartheta T\right)^2 = E_\vartheta T^2 - e^{-2\vartheta}$$

$$= e^{-2\vartheta}\left(e^{\vartheta/n} - 1\right) \geq e^{-2\vartheta}\frac{\vartheta}{n}$$

Since $g'(\vartheta) = -e^{-\vartheta}$

$$\frac{g'(\vartheta)^2}{n I_1(\vartheta)} = \frac{\vartheta e^{-2\vartheta}}{n}$$

$\Rightarrow$ UMVU estimator does not reach the CRLB.