THE CRAMER-RAO LOWER BOUND

The CRLB is a lower bound on the variance of any unbiased estimator of a parameter $\underline{\theta}$.
It is useful in many ways:

1. If $\hat{\underline{\theta}}$ achieves the CRLB for all $\underline{\theta} \in \Theta$, the $\hat{\underline{\theta}}$ is a MVUE.

2. The CRLB provides a benchmark against which we can compare the performance of any unbiased estimator. We're doing well if our estimator is "close" to the CRLB

3. The CRLB allows us to rule out impossible estimators. We know it is impossible to find an estimator that beats the CRLB. This is useful in feasibility studies.

4. The theory behind the CRLB tells us precisely when the bound is achievable.

$\boxed{\text{CRLB: Scalar Parameter}}$

Consider $\underline{X} \sim f(\underline{x};\theta)$ where $\theta \in \Theta \subseteq \mathbb{R}$. $\underline{X}$ may be continuous or discrete. The function

$$\frac{\partial}{\partial\theta} \ln f(\underline{x};\theta)$$

is called the <u>score function</u>. The quantity

$$I(\theta) = E\left\{\left(\frac{\partial}{\partial\theta} \ln f(\underline{X};\theta)\right)^2\right\}$$

is called the <u>Fisher information</u>.

<u>Theorem</u> Suppose $I(\theta)$ exists and that f satisfies certain regularity conditions. Then the variance of any unbiased estimator $\hat{\theta}$ satisfies

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)},$$

where $I(\theta)$ is evaluated at the true parameter.

Furthermore, the bound holds with equality iff

$$\frac{\partial}{\partial\theta} \ln f(\underline{x};\theta) = I(\theta)\cdot(\hat{\theta}(\underline{x}) - \theta) \quad \forall \underline{x}$$

Remarks

1. If equality holds, we say $\hat{\theta}$ is efficient.
2. If $\hat{\theta}$ is efficient, then clearly $\hat{\theta}$ is a MVUE. However, not all MVUEs are efficient. Thus, efficient estimators do not always exist.
3. When viewed as a function of θ, $f(\underline{x};\theta)$ is called the likelihood of θ, and $\ln f(\underline{x};\theta)$ the log-likelihood.
4. Using integration by parts, it can be shown that
$$I(\theta) = -E\left\{\frac{\partial^2}{\partial\theta^2}\ln f(\underline{X};\theta)\right\}$$
This expression is often easier to compute.
5. The CRLB says nothing about the performance of biased estimators. The variance of a biased estimator may very well be less than the CRLB.

<u>Example</u> Suppose $\underline{x} = [x_1, \ldots, x_N]^T$ where

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2), \qquad i = 1, \ldots, N$$

with $\theta = \mu$ (σ^2 known).

$$\ln f(\underline{x}; \mu) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2$$

$$\frac{\partial}{\partial\mu} \ln f(\underline{x}; \mu) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$$

$$\frac{\partial^2}{\partial\mu} \ln f(\underline{x}; \mu) =$$

$$I(\mu) = -E\left\{-\frac{N}{\sigma^2}\right\} = \frac{N}{\sigma^2}$$

$\Longrightarrow$ $Var(\hat{\mu}) \geq \frac{\sigma^2}{N}$ for any unbiased estimator $\hat{\mu}$ of μ

Also, we have

$$\frac{\partial}{\partial\mu} \ln f(\underline{x}; \mu) = \frac{1}{\sigma^2} \sum (x_i - \mu)$$

$$= \frac{N}{\sigma^2} \left(\frac{1}{N} \sum_{i=1}^{N} x_i - \mu \right)$$

$$= I(\mu) \cdot (\hat{\mu}(\underline{x}) - \mu)$$

$\Longrightarrow$ $\hat{\mu}(\underline{x}) = \frac{1}{N} \sum x_i$ is efficient.

Exercise Consider the correlated bivariate Gaussian

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

where ρ is known. Find the CRLB. Hint:

$$\log f(\underline{x}\,;\mu) = \frac{-1}{1+\rho}\left(\mu^2 - \mu(x_1 + x_2)\right) + C$$

Does an efficient estimator exist?

Solution |

$$\frac{\partial}{\partial \mu} \ln f(\underline{x};\mu) = -\frac{(2\mu - (x_1 + x_2))}{1+\rho}$$

$$\frac{\partial^2}{\partial \mu^2} \ln f(\underline{x};\mu) = -\frac{2}{1+\rho}$$

$$I(\mu) = -E\left\{-\frac{2}{1+\rho}\right\} = \frac{2}{1+\rho}$$

$$\Rightarrow \text{CRLB} = \frac{1+\rho}{2}$$

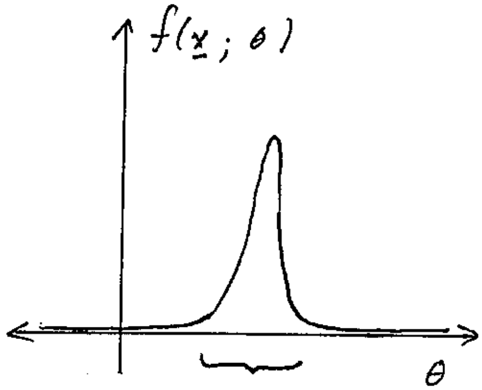
$\boxed{\text{Interpretation?}}$

$$\frac{\partial \ln f(\underline{x};\mu)}{\partial \mu} = \frac{2}{1+\rho}\left(\frac{x_1 + x_2}{2} - \mu\right)$$

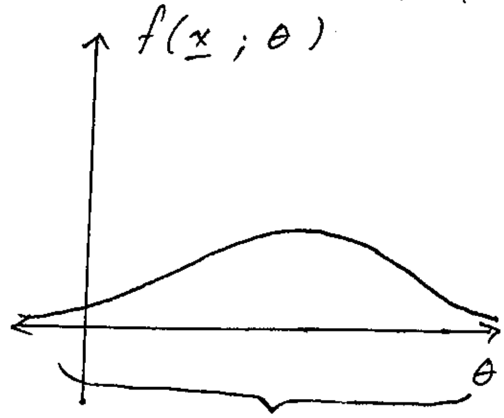
efficient

Fisher Information and Average Curvature

Skip in lecture

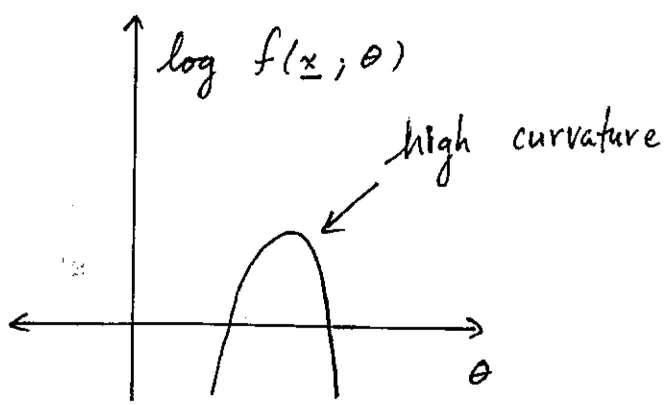


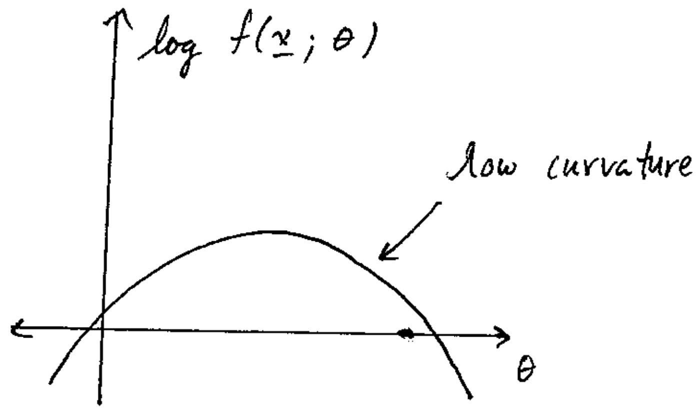
highly likely that θ is in this range



harder to pinpoint θ

The operator $-\dfrac{\partial^2}{\partial\theta^2}$ measures curvature





So $I(\theta)$ reflects the average curvature of the log-likelihood $\log f(\underline{x};\theta)$

<u>Conclusion</u>: θ is easy to estimate

$\Longleftrightarrow$ $f(\underline{x};\theta)$ is "peaky" near θ (on average)

$\Longleftrightarrow$ $\log f(\underline{x};\theta)$ has high curvature at θ (on average)

$\Longleftrightarrow$ $I(\theta) = -E\left\{\dfrac{\partial^2}{\partial\theta^2}\log f(\underline{X};\theta)\right\}$ is large

$\Longleftrightarrow$ CRLB is small

Proof of CRLB: Scalar Case

First observe

$$E\left\{\frac{\partial}{\partial\theta}\ln f(\underline{X};\theta)\right\} = \int\left(\frac{\partial}{\partial\theta}\ln f(\underline{x};\theta)\right) f(\underline{x};\theta)\,d\underline{x}$$

$$= \int \frac{\left(\frac{\partial}{\partial\theta} f(\underline{x};\theta)\right)}{f(\underline{x};\theta)} \cdot f(\underline{x};\theta)\,d\underline{x}$$

$$= \int \frac{\partial}{\partial\theta} f(\underline{x};\theta)\,d\underline{x}$$

$$= \frac{\partial}{\partial\theta}\int f(\underline{x};\theta)\,d\underline{x}$$

$$= \frac{\partial}{\partial\theta}\{1\}$$

$$= 0$$

This shows

$$E\left\{\frac{\partial}{\partial\theta}\ln f(\underline{\mathbf{X}};\theta)\right\} = 0 \qquad (*)$$

that is, the score function is zero on average.

Next, since $\hat{\theta}$ is unbiased,

$$\theta = \int \hat{\theta}(\underline{x})\, f(\underline{x};\theta)\, d\underline{x}$$

Differentiating both sides with respect to θ,

$$
\begin{aligned}
1 &= \frac{\partial}{\partial\theta} \int \hat{\theta}(\underline{x})\, f(\underline{x};\theta)\, d\underline{x} \\
&= \int \hat{\theta}(\underline{x}) \frac{\partial}{\partial\theta} f(\underline{x};\theta)\, dx \\
&= \int \hat{\theta}(\underline{x}) \cdot \frac{\left(\frac{\partial}{\partial\theta} f(\underline{x};\theta)\right)}{f(\underline{x};\theta)}\, f(\underline{x};\theta)\, d\underline{x} \\
&= \int \hat{\theta}(\underline{x}) \left(\frac{\partial}{\partial\theta} \ln f(\underline{x};\theta)\right) f(\underline{x};\theta)\, dx \\
&= \int \left(\hat{\theta}(\underline{x}) - \theta\right) \frac{\partial}{\partial\theta} \ln f(\underline{x};\theta)\, f(\underline{x};\theta)\, d\underline{x} \\
&= E\{Y \cdot Z\}
\end{aligned}
$$

by (*)

where

$$Y = \hat{\theta}(\underline{X}) - \theta$$

$$Z = \frac{\partial}{\partial\theta} \ln f(\underline{X};\theta).$$

Finally, the CRLB follows from the Cauchy-Schwarz inequality:

$$
\begin{aligned}
1 &= E\{Y \cdot Z\} \\
&= (E\{Y \cdot Z\})^2 \\
&\leq E\{Y^2\} \cdot E\{Z^2\} \\
&= E\{(\hat{\theta}(\underline{X}) - \theta)^2\} \cdot E\left\{\left(\frac{\partial}{\partial \theta} \ln f(\underline{X}; \theta)\right)^2\right\} \\
&= \mathrm{Var}(\hat{\theta}) \cdot I(\theta)
\end{aligned}
$$

$$
\Longrightarrow \quad \mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}
$$

The "regularity conditions" are essentially that it is valid to exchange the derivative and integral, as was done twice in the above. This should be carefully checked when using the CRLB. See Poor or Scharf for additional discussion.

Equality holds in the Cauchy-Schwarz inequality iff $\exists\, k(\theta)$ (a constant not depending on $\underline{x}$) such that

$$\frac{\partial \log f(\underline{x}\,;\theta)}{\partial \theta} = k(\theta)\left(\hat{\theta}(\underline{x}) - \theta\right) \qquad \forall\, \underline{x} \in \mathcal{X}$$

Taking the derivative w.r.t θ of both sides

$$\frac{\partial^2}{\partial \theta^2} \log f(\underline{x}\,;\theta) = -k(\theta) + k'(\theta)\cdot\left(\hat{\theta}(\underline{x}) - \theta\right),$$

and taking $-E\{\cdot\}$ we get

$$I(\theta) = -E\left\{\frac{\partial^2}{\partial \theta^2} \log f(\underline{x}\,;\theta)\right\}$$

$$= k(\theta) - k'(\theta)\, \underbrace{E\left\{\hat{\theta}(\underline{X}) - \theta\right\}}_{=\,0}$$

$$= k(\theta). \qquad \blacksquare$$

CRLB: Vector Parameter

The vector CRLB has the form

$$\mathrm{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq I(\underline{\theta})^{-1}$$

where

$$\mathrm{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) = E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\right\}$$

$$= \begin{bmatrix} \mathrm{Var}(\hat{\theta}_1) & \mathrm{Cov}(\hat{\theta}_1, \theta_2) & \cdots & \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_p) \\ \mathrm{Cov}(\hat{\theta}_2, \hat{\theta}_1) & \mathrm{Var}(\hat{\theta}_2) & & \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(\hat{\theta}_p, \hat{\theta}_1) & & & \mathrm{Var}(\hat{\theta}_p) \end{bmatrix}$$

is the <u>covariance matrix</u> of $\hat{\underline{\theta}}$ and

$$I(\underline{\theta}) = E\left\{\left(\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}; \underline{\theta})\right)\left(\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}; \underline{\theta})\right)^T\right\}$$

is the <u>Fisher information matrix</u> of $\underline{\theta}$, and

$$\mathrm{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq I(\underline{\theta})^{-1}$$

means

$$\mathrm{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) - I(\underline{\theta})^{-1} \text{ is positive semi-definite.}$$

Recall that if $\phi : \mathbb{R}^p \longrightarrow \mathbb{R}$ then

$$\frac{\partial \phi}{\partial \underline{\theta}} = \left[\frac{\partial \phi}{\partial \theta_1} \cdots \frac{\partial \phi}{\partial \theta_p} \right]^T =: \nabla_{\underline{\theta}}\, \phi$$

alternate notation

Analogous to the scalar case, it can be shown that

$$I(\underline{\theta}) = E\left\{ \left(\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}; \theta) \right) \left(\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}; \underline{\theta}) \right)^T \right\}$$

$$= -E\left\{ \frac{\partial^2}{\partial \underline{\theta}\, \partial \underline{\theta}^T} \log f(\underline{X}; \underline{\theta}) \right\}$$

where

$$\frac{\partial \phi}{\partial \underline{\theta}^T} = \left[\frac{\partial \phi}{\partial \theta_1} \cdots \frac{\partial \phi}{\partial \theta_p} \right] = \left(\frac{\partial \phi}{\partial \underline{\theta}} \right)^T$$

and

$$\frac{\partial^2 \phi}{\partial \underline{\theta}\, \partial \underline{\theta}^T} = \begin{bmatrix} \frac{\partial^2 \phi}{\partial \theta_1^2} & \frac{\partial^2 \phi}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \phi}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 \phi}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \phi}{\partial \theta_2^2} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 \phi}{\partial \theta_p \partial \theta_1} & & \cdots & \frac{\partial^2 \phi}{\partial \theta_p^2} \end{bmatrix} =: \nabla_{\underline{\theta}}^2 \phi$$

Theorem

Suppose $\underline{I}(\underline{\theta})$ exists and that f satisfies certain regularity conditions. If $\hat{\underline{\theta}}$ is an unbiased estimator of $\underline{\theta}$, then

$$\mathrm{Cov}(\hat{\underline{\theta}}) \geq \underline{I}(\underline{\theta})^{-1}$$

with equality iff

$$\frac{\partial}{\partial \underline{\theta}} \ln f(\underline{x}; \underline{\theta}) = \underline{I}(\underline{\theta}) \cdot (\hat{\underline{\theta}}(\underline{x}) - \underline{\theta}) \quad \forall \underline{x}$$

Proof For the most part, the proof generalizes the proof of the scalar case, although some new techniques are necessary. See Hero or Scharf for details.

If A and B are symmetric matrices and $A \geq B$, then $a_{ii} \geq b_{ii}$ $\forall i$. This follows by taking $\underline{z}_i = [0 \cdots 0 \; 1 \; 0 \cdots 0]^T$ and noting

$$\begin{array}{c} \uparrow \\ i\text{th position} \end{array}$$

$$0 \leq \underline{z}_i^T (A - B) \underline{z}$$

$$= a_{ii} - b_{ii}.$$

Therefore we have the following

Corollary | Under the assumptions of the CRLB, if $\underline{\hat{\theta}}$ is unbiased then

$$\text{Var}(\hat{\theta}_i) \geq [I(\underline{\theta})]^{-1}_{ii}$$

Thus, the vector CRLB implies scalar lower bounds on each component of $\underline{\hat{\theta}}$.

Furthermore, if $\text{Cov}_{\underline{\theta}}(\underline{\hat{\theta}}) = I(\underline{\theta})^{-1}$ $\forall \underline{\theta}$, then $\underline{\hat{\theta}}$ is a MVUE because

$$\begin{aligned} \text{Var}_{\underline{\theta}}(\underline{\hat{\theta}}) &= E\left\{(\underline{\hat{\theta}} - \underline{\theta})^T(\underline{\hat{\theta}} - \underline{\theta})\right\} \\ &= E\left\{\sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2\right\} = \sum_{i=1}^{N} E\left\{(\hat{\theta}_i - \theta_i)^2\right\} \\ &= \sum_{i=1}^{N} \text{Var}(\hat{\theta}_i) \end{aligned}$$

is minimized.

Exercise Suppose $\underline{x} = [x_1 \cdots x_N]^T$ where

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2).$$

Find the CRLB for $\underline{\theta} = [\mu \;\; \sigma^2]^T$.

Note: $$\log f(\underline{x}; \underline{\theta}) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2$$

Solution 1

$$\frac{\partial \log f(\underline{x};\underline{\theta})}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu)$$

$$\frac{\partial \log f(\underline{x};\underline{\theta})}{\partial(\sigma^2)} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\frac{\partial^2 \log f(\underline{x};\underline{\theta})}{\partial \mu^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial^2 \log f(\underline{x};\underline{\theta})}{\partial(\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\frac{\partial^2 \log f(\underline{x};\underline{\theta})}{\partial \mu\,\partial(\sigma^2)} = \frac{-1}{\sigma^4}\sum_{i=1}^{N}(x_i-\mu)$$

$$I(\underline{\theta}) = -E\begin{bmatrix} -\frac{N}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{i=1}^{N}(X_i-\mu) \\ -\frac{1}{\sigma^4}\sum_{i=1}^{N}(X_i-\mu) & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(X_i-\mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

$$\Longrightarrow \quad Var(\hat{\mu}) \geq \frac{\sigma^2}{N}$$

$$Var(\widehat{\sigma^2}) \geq \frac{2\sigma^4}{N}$$

$\boxed{\text{Summary}}$

- CRLB = lower bound on variance of any unbiased estimator
- Bound given by Fisher Information (matrix)
- Score function = $\frac{\partial}{\partial \theta} \ln f(\underline{x}; \underline{\theta})$ determines condition for equality
- Proof: application of Cauchy-Schwarz inequality

The Cauchy Schwarz Inequality

An <u>inner product space</u> is a vector space equipped with an inner product $\langle \cdot, \cdot \rangle$, which is an operator $V \to \mathbb{C}$ that satisfies

- $\langle u, v \rangle = \langle v, u \rangle^*$
- $\langle a_1 u_1 + a_2 u_2, v \rangle = a_1 \langle u_1, v \rangle + a_2 \langle u_2, v \rangle$
- $\langle u, u \rangle \geq 0$ with equality iff $u = 0$

Grammians / Gram Matrices

Let V be an IPS, and $v_1, \dots, v_N \in V$. Set

$$G = \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_N \rangle \\ \langle v_2, v_1 \rangle & \cdots & \langle v_2, v_N \rangle \\ \vdots & \ddots & \\ \langle v_N, v_1 \rangle & & \langle v_N, v_N \rangle \end{bmatrix} \in \mathbb{C}^{N \times N}$$

<u>Theorem</u> G is PSD, and G is PD $\iff$ $v_1, \dots, v_N$ are LI (linearly independent)

Proof: Let $\underline{x} = [x_1 \cdots x_N]^T \in \mathbb{C}^N$. Then

$$\underline{x}^H G \underline{x} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j^* \langle v_i, v_j \rangle$$

$$= \sum_{i=1}^{N} x_i \langle v_i, \sum_{j=1}^{N} x_j v_j \rangle \qquad \circledast$$

$$= \langle \sum_{i=1}^{N} x_i v_i, \sum_{j=1}^{N} x_j v_j \rangle$$

$$\geq 0$$

with equality iff $\sum x_i v_i = 0$.

To see $\circledast$, note that

$$x_j^* \langle v_i, v_j \rangle = x_j^* \langle v_j, v_i \rangle^* = (x_j \langle v_j, v_i \rangle)^*$$

$$= \langle x_j v_j, v_i \rangle^* = \langle v_i, x_j v_j \rangle. \qquad \square$$

Corollary: $|\langle v_1, v_2 \rangle| \leq \sqrt{\langle v_1, v_1 \rangle \cdot \langle v_2, v_2 \rangle}$ with equality iff v_1 and v_2 are linearly dependant.

Proof: Consider

$$G = \begin{bmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle \end{bmatrix}$$

G is PSD $\Rightarrow$ $\det G \geq 0$ (because $\det G$ $=$ product of eigenvalues of G). Now

$$\det G = \langle v_1, v_1 \rangle \langle v_2, v_2 \rangle - \underbrace{\langle v_1, v_2 \rangle \cdot \langle v_2, v_1 \rangle}_{= |\langle v_1, v_2 \rangle|^2}$$

Equality holds $\Leftrightarrow$ G is not PD $\Leftrightarrow$ v_1, v_2 are linearly dependent. $\square$.

<u>Application</u>

$V = \{$zero mean real-valued random variables with finite variance$\}$

$\langle Y, Z \rangle = E\{YZ\}$ $\leftarrow$ verify this is a valid inner prod.

$\langle Y, Y \rangle = \mathrm{Var}\{Y\}$

Therefore $|E\{Y \cdot Z\}| \leq \sqrt{\mathrm{Var}(Y) \cdot \mathrm{Var}(Z)}$.