Cramer-Rao Lower Bound

Nart Shawash
ELEE 5810

April 19, 2021

Table of contents

CRLB applies to PDFs satisfying *regularity condition*

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial \ln p(\mathrm{x};\theta)}{\partial\theta}\right] &= \mathbb{E}\left[\frac{1}{p(\mathrm{x};\theta)}\frac{\partial p(\mathrm{x};\theta)}{\partial\theta}\right] \\
&= \int p(\mathrm{x};\theta)\left[\frac{1}{p(\mathrm{x};\theta)}\frac{\partial p(\mathrm{x};\theta)}{\partial\theta}\right] d\mathrm{x} \\
&= \int \frac{\partial p(\mathrm{x};\theta)}{\partial\theta} d\mathrm{x} \\
&= \frac{\partial}{\partial\theta}\left[\underbrace{\int p(\mathrm{x};\theta)\, d\mathrm{x}}_{1}\right] \\
&= 0
\end{aligned}
$$

Cramer-Rao Lower Bound aka Cramer-Rao Inequality

Theorem

If the PDF $p(\mathrm{x};\theta)$ satisfies the regularity condition for all θ, then the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$var(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};\theta)}{\partial\theta^2}\right]}$$

Furthermore:

$$\boxed{\text{Can find } \underbrace{\text{MVU estimator}}_{\hat{\theta}=g(\mathrm{x})} \forall\theta} \overset{iff}{\Longleftrightarrow} \boxed{\frac{\partial \ln p(\mathrm{x};\theta)}{\partial\theta} = I(\theta)(g(\mathrm{x})-\theta)}$$

Where $\frac{1}{I(\theta)} = \sigma^2$ *is the minimum variance.*

Example 3.3: DC in AWGN

Consider N observations $\boxed{x[n] = A + w[n]}$. $w[n] \sim \mathcal{N}(0, \sigma^2)$.

1. Note that $p(\mathrm{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \cdot e^{-\frac{\sum_{n=0}^{N-1}(x[n]-A)^2}{2\sigma^2}}$
2. $\frac{\partial \ln p(\mathrm{x};A)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n] - A) = \frac{N}{\sigma^2}(\bar{x} - A) \stackrel{?}{=} I(A)(g(\mathrm{x}) - A)$
3. $\frac{\partial^2 \ln p(\mathrm{x};A)}{\partial A^2} = -\frac{N}{\sigma^2}$, and $\boxed{-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};A)}{\partial A^2}\right] = \frac{N}{\sigma^2} \stackrel{?}{=} I(A)}$
4. CRLB says

$$var(\hat{A}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};A)}{\partial A^2}\right]} = \frac{\sigma^2}{N}$$

5. Hence the bound is attained, with $g(\mathrm{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \bar{x}$

Important Identity, Problem 3.8

$$
\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};\theta)}{\partial\theta^2}\right] &= -\mathbb{E}\left[\frac{\partial}{\partial\theta}\left(\frac{1}{p(\mathrm{x};\theta)}\frac{\partial p(\mathrm{x};\theta)}{\partial\theta}\right)\right] \\
&= -\mathbb{E}\left[\frac{1}{p(\mathrm{x};\theta)}\frac{\partial^2 p(\mathrm{x};\theta)}{\partial\theta^2} - \frac{1}{p^2(\mathrm{x};\theta)}\left(\frac{\partial p(\mathrm{x};\theta)}{\partial\theta}\right)^2\right] \\
&= -\int \frac{\partial^2 p(\mathrm{x};\theta)}{\partial\theta^2} d\mathrm{x} + \int \frac{p(\mathrm{x};\theta)}{p^2(\mathrm{x};\theta)}\left(\frac{\partial p(\mathrm{x};\theta)}{\partial\theta}\right)^2 d\mathrm{x} \\
&= -\frac{\partial^2}{\partial\theta^2}\underbrace{\left[\int p(\mathrm{x};\theta) d\mathrm{x}\right]}_{1} + \int p(\mathrm{x};\theta)\left(\underbrace{\frac{1}{p(\mathrm{x};\theta)}\frac{\partial p(\mathrm{x};\theta)}{\partial\theta}}_{\frac{\partial \ln p(\mathrm{x};\theta)}{\partial\theta}}\right)^2 d\mathrm{x} \\
&= \mathbb{E}\left[\left(\frac{\partial \ln p(\mathrm{x};\theta)}{\partial\theta}\right)^2\right] \geq 0 \;\; \forall\theta
\end{aligned}
$$

Example 3.4 - Phase Estimation

Want to estimate phase ϕ of a sinusoid in AWGN $w[n] \sim \mathcal{N}(0, \sigma^2)$

$$x[n] = A\cos(2\pi f_0 n + \phi) + w[n] \quad n = 0, 1, \ldots, N-1$$

here both A and $0 \leq f_0 \leq \frac{1}{2}$ are known.

$$p(\mathrm{x}; \phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \cdot e^{-\frac{\sum_{n=0}^{N-1}[x[n] - A\cos(2\pi f_0 n + \phi)]^2}{2\sigma^2}}$$

$$\frac{\partial \ln p(\mathrm{x}; \phi)}{\partial \phi} = -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}[x[n] - \underbrace{A\cos(2\pi f_0 n + \phi)]\, A\sin(2\pi f_0 n + \phi)}_{\sin\alpha\cos\beta = \frac{1}{2}(\sin(\alpha+\beta) - \sin(\alpha-\beta))}$$

$$= -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[x[n]\sin(2\pi f_0 n + \phi) - \frac{A}{2}\sin(4\pi f_0 n + 2\phi)\right]$$

Example 3.4 - Phase Estimation (continued)

$$\frac{\partial^2 \ln p(\mathrm{x};\phi)}{\partial \phi^2} = -\frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[x[n]\cos(2\pi f_0 n+\phi) - A\cos(4\pi f_0 n + 2\phi)\right]$$

substitute $x[n] = A\cos\left(2\pi f_0 n + \phi\right) + w[n]$

$$-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};\phi)}{\partial \phi^2}\right] = \frac{A}{\sigma^2}\sum_{n=0}^{N-1}\left[A\underbrace{\cos^2(2\pi f_0 n+\phi)}_{\cos^2\theta=\frac{1}{2}[1+\cos 2\theta]} - A\cos(4\pi f_0 n + 2\phi)\right]$$

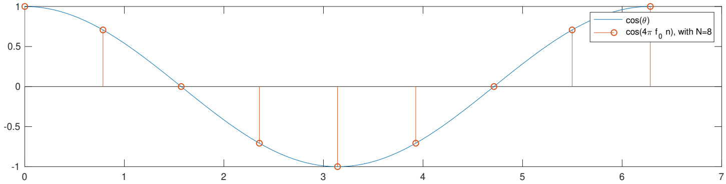
$$= \frac{A^2}{\sigma^2}\sum_{n=0}^{N-1}\left[\frac{1}{2} + \underbrace{\frac{1}{2}\cos(4\pi f_0 n+2\phi) - \cos(4\pi f_0 n+2\phi)}_{\frac{1}{N}\sum_{n=0}^{N-1}\cos(4\pi f_0 n+2\phi)\approx 0,\ \text{Problem 3.7}}\right]$$

$$\approx \frac{NA^2}{2\sigma^2} \qquad \text{provided } f_0 \text{ not near 0 or } \frac{1}{2}$$

Example 3.4 - Phase Estimation (continued)

For what f_0 does $\boxed{\frac{1}{N}\sum_{n=0}^{N-1}\cos(4\pi f_0 n + 2\phi) \approx 0}$ hold?

1. Need Euler's identity $\boxed{Re[e^{i\theta}] = Re[\cos\theta + i\sin\theta] = \cos\theta}$
2. Need finite geometric sum $\boxed{\sum_{n=0}^{N-1} r^n = \frac{1-r^N}{1-r}, \text{ for } r \neq 1}$
3. Note that $0 < f_0 < \frac{1}{2}$, (otherwise $r = 1$, see below).

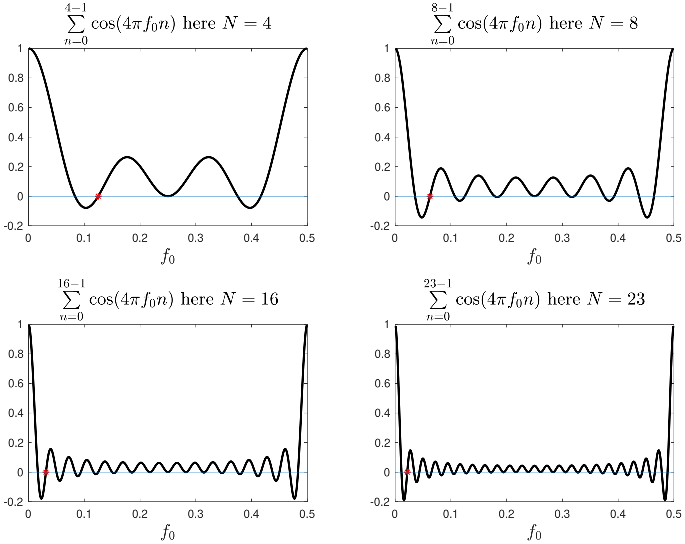


Example 3.4 - Phase Estimation (continued)

$$\begin{aligned}
\frac{1}{N}\sum_{n=0}^{N-1}\cos(4\pi f_0 n) &= \frac{1}{N} Re\left(\sum_{n=0}^{N-1}\left[e^{i4\pi f_0}\right]^n\right) \\
&= \frac{1}{N} Re\left(\frac{1-e^{i4\pi f_0 N}}{1-e^{i4\pi f_0}}\cdot\frac{1-e^{-i4\pi f_0}}{1-e^{-i4\pi f_0}}\right) \\
&= \frac{1}{2N}\cdot\frac{Re\left(1-e^{-i4\pi f_0}-e^{i4\pi f_0 N}+e^{i4\pi f_0(N-1)}\right)}{1-\cos(4\pi f_0)} \\
&= \frac{1}{2N}\cdot\frac{1-\cos(4\pi f_0)\overbrace{-\cos(4\pi f_0 N)+\cos(4\pi f_0(N-1))}^{\overset{?}{\approx}0}}{1-\cos(4\pi f_0)} \\
&\approx \frac{1}{2N}\approx 0, \quad \text{for large } N
\end{aligned}$$

You repeat this for $\frac{1}{N}\sum_{n=0}^{N-1}\cos(4\pi f_0 n + 2\phi)\approx 0$

Example 3.4 - Phase Estimation (continued)



Example 3.4 - Phase Estimation (continued)

Back to

$$var(\hat{\phi}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};\phi)}{\partial \phi^2}\right]} \approx \frac{2\sigma^2}{NA^2}$$

Note that the **Furthermeore, iff** condition(s) of CRLB are not satisfied

$$\frac{\partial \ln p(\mathrm{x};\phi)}{\partial \phi} = I(\phi)(\hat{\phi} - \phi)$$

$$\frac{\partial^2 \ln p(\mathrm{x};\phi)}{\partial \phi^2} = \frac{\partial I(\phi)}{\partial \phi}(\hat{\phi} - \phi) - I(\phi)$$

$$-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};\phi)}{\partial \phi^2}\right] = -\frac{\partial I(\phi)}{\partial \phi}\underbrace{(\mathbb{E}(\hat{\phi}) - \phi)}_{\text{not unbiased}} + I(\phi)$$

So we do not (yet) know how to find MVU estimator. ☹

General CRLB for signals in AWGN $x[n] = s[n;\theta] + w[n]$

$s[n;\theta]$ a deterministic signal with unknown parameter θ

$$p(\mathrm{x};\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \cdot e^{-\frac{1}{2\sigma^2}\sum\limits_{n=0}^{N-1}(x[n]-s[n;\theta])^2}$$

$$\frac{\partial \ln p(\mathrm{x};\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n;\theta])\frac{\partial s[n;\theta]}{\partial\theta}$$

$$\frac{\partial^2 \ln p(\mathrm{x};\theta)}{\partial\theta^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left[(x[n]-s[n;\theta])\frac{\partial^2 s[n;\theta]}{\partial\theta^2} - \left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2\right]$$

$$\mathbb{E}\left[\frac{\partial^2 \ln p(\mathrm{x};\theta)}{\partial\theta^2}\right] = -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2,$$ so CRLB says

$$\mathrm{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2},$$ note the kind of dependence on θ

Example 3.5 - Sinusoid frequency Estimation

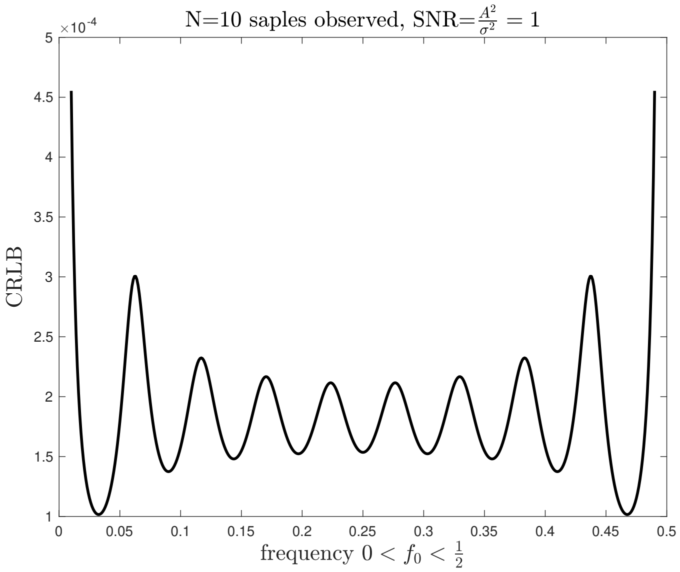
In the model above, let the signal be

$$s[n; f_0] = A\cos(2\pi f_0 n + \phi), \quad 0 < f_0 < \frac{1}{2}$$

Here, $\{A, \phi\}$ are known, but f_0 is the parameter to be estimated. Direct substitution (last Eq. on previous slide) gives

$$\text{var}(\hat{f}_0) \geq \frac{\sigma^2}{A^2 \sum_{n=0}^{N-1} [2\pi n \sin(2\pi f_0 n + \phi)]^2}$$

Example 3.5 - Sinusoid frequency Estimation (continued)



Example 3.5 - Sinusoid frequency Estimation (continued)

Note that in this example

$$\lim_{f_0 \to 0} \text{CRLB} = \lim_{f_0 \to 0} \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]}$$

$$= \lim_{f_0 \to 0} \frac{1}{\mathbb{E}\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]}$$

$$= \lim_{f_0 \to 0} \frac{\sigma^2}{A^2 \sum_{n=0}^{N-1} [2\pi n \sin(2\pi f_0 n + \phi)]^2}$$

$$= \infty$$

That is for f_0 close to 0, slight change in frequency will not alter the signal significantly; this is a weak dependence on $\theta = f_0$.

§3.6 Transformation of Parameters

- We might not be interested in the sign of A, but in the signal power $\alpha = g(A) = A^2$.
- If we know CRLB for A, we can obtain CRLB for $\alpha = g(\theta)$ (see Appendix 3A) using

$$\text{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-\mathbb{E}\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]}$$

- In the"signal power" case: $\alpha = g(A) = A^2$ and

$$\text{var}(\widehat{A^2}) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

§3.6 Transformation of Parameters (continued)

- Recall that sample mean estimator $\bar{x}$ is efficient for A.
- Is $\bar{x}^2$ an efficient estimator for A^2? That is $\boxed{\mathbb{E}[\bar{x}^2] \stackrel{?}{=} A^2}$?
- First recall that $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$, and $\mathbb{E}[\bar{x}] = A$, $\text{var}(\bar{x}) = \frac{\sigma^2}{N}$

$$\begin{aligned}
\text{var}(\bar{x}) &= \mathbb{E}\left[(\bar{x} - A)^2\right] \\
&= \mathbb{E}\left[\bar{x}^2 - 2A\bar{x} + A^2\right] \\
&= \mathbb{E}\left[\bar{x}^2\right] - 2A\mathbb{E}[\bar{x}] + A^2 \\
&= \mathbb{E}[\bar{x}^2] - A^2 = \mathbb{E}[\bar{x}^2] - \mathbb{E}^2[\bar{x}] \\
\mathbb{E}[\bar{x}^2] &= A^2 + \text{var}(\bar{x}) \neq A^2 \quad \text{nonlinearity destroyed efficiency} \\
&= A^2 + \underbrace{\frac{\sigma^2}{N}}_{\text{bias}}, \text{ but } \lim_{N\to\infty} \text{bias} = 0, \quad \text{long data record}
\end{aligned}$$

§3.6 Transformation of Parameters (continued)

- But what if $g(\theta)$ is affine? That is $g(\theta) = a\theta + b$.
- First:
  1. $\mathbb{E}[g(\hat{\theta})] = \mathbb{E}[a\hat{\theta} + b] = a\mathbb{E}[\hat{\theta}] + b = a\theta + b$
  2. $\text{var}(\widehat{g(\theta)}) = \text{var}(a\hat{\theta} + b) = a^2 \text{var}(\hat{\theta})$
- Next, CRLB

$$\text{var}(\widehat{g(\theta)}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-\mathbb{E}\left[\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right]} = \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{I(\theta)} = \left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \text{var}(\hat{\theta})$$
$$= a^2 \text{var}(\hat{\theta})$$

- Hence the CRLB is achieved. ☺

§3.6 Transformation of Parameters (continued)

- Let's revisit $\bar{x}^2$ as an estimator for A^2.
- How bad is $\text{var}(\bar{x}^2) = \mathbb{E}[\bar{x}^4] - \mathbb{E}^2[\bar{x}^2]$?
- To answer this, we need the following:

if $\zeta \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(\zeta-\mu)^2}{2\sigma^2}}$, then

$$\mathbb{E}[\zeta^2] = \int_{-\infty}^{\infty} \zeta^2 \, \mathcal{N}(\mu, \sigma^2) \, d\zeta = \mu^2 + \sigma^2$$

$$\mathbb{E}[\zeta^4] = \int_{-\infty}^{\infty} \zeta^4 \, \mathcal{N}(\mu, \sigma^2) \, d\zeta = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

$$\text{var}(\zeta^2) = \mathbb{E}[\zeta^4] - \mathbb{E}[\zeta^2] = 4\mu^2\sigma^2 + 2\sigma^4$$

- Hence in our problem asymptotic efficiency is achieved

$$\text{var}(\bar{x}^2) = \frac{4A^2\sigma^2}{N} + \underbrace{\frac{2\sigma^4}{N^2}}_{\text{decays fast}}$$

§3.6 Transformation of Parameters (continued)

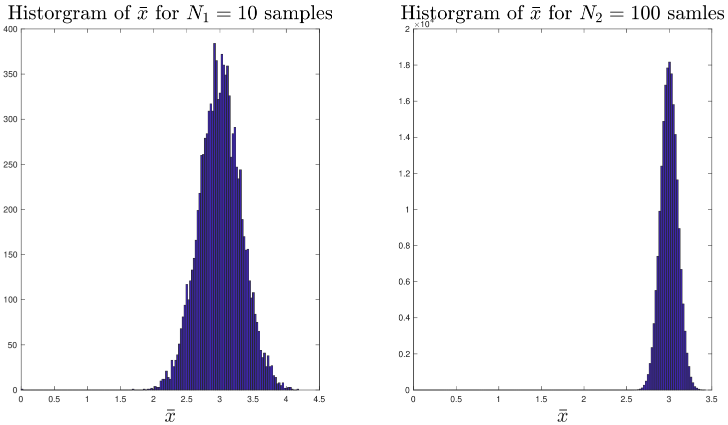
Finally,

1. As N increases, the PDF of $\bar{x}$ becomes more concentrated about the mean A.
2. Therefore, the observed values of $\hat{x}$ lie in smaller region around A.
3. Hence, it makes sense to linearize the nonlinear transformation $g(\hat{x}) = \hat{x}^2$ around $\hat{x} = A$.

$$\begin{aligned} g(\bar{x}) &\approx g(A) + \frac{d}{d\bar{x}} [g(\bar{x})]\Big|_{\bar{x}=A} (\bar{x} - A) \\ &= A^2 + 2A(\bar{x} - A) \end{aligned}$$

4. It follows that:
   - $\mathbb{E}[g(\bar{x})] = g(A) = A^2$
   - $\text{var}(g(\bar{x})) = \left[\frac{d\,g(A)}{dA}\right]^2 \cdot \text{var}(\bar{x}) = (2A)^2 \cdot \frac{\sigma^2}{N} = \frac{4A^2\sigma^2}{N}$

§3.6 Transformation of Parameters (continued)



```
N1=10; % length of data record
xh1=0;  A=3;      % DC in AWGN
for i=1:200000
    xh1=[xh1 mean(A+randn(1,N1))];
end
xh1=xh1(1,1:size(xh1,2)-1);
subplot(1,2,1); hist(xh,150); xlabel('$\bar{x}$');
title('Historgram of $\bar{x}$ for $N_1=10$ samples');

N2=100; % length of data record
xh2=0;  A=3;      % DC in AWGN
for i=1:200000
    xh2=[xh2 mean(A+randn(1,N2))];
end
xh2=xh2(1,1:size(xh2,2)-1);
subplot(1,2,2); hist(xh2,150); xlabel('$\bar{x}$');
title('Historgram of $\bar{x}$ for $N_2=100$ samles');
```

§3.6 Transformation of Parameters (continued)

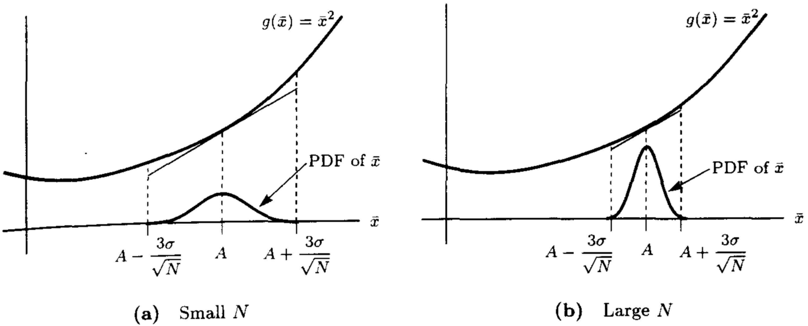


Figure: Linearization of nonlinear transformation.