Iterative Least Squares and M-Estimation for ARMA Processes With Infinite Variance

Richard A. Davis
Colorado State University
rdavis@stat.colostate.edu

- Something old
  - sample acf
  - estimation
- Something new
  - M-estimation
  - LAD-estimation
  - linearized estimates
- Something blue
  - bootstrapping
  - MLE

Sample ACF .

Let $\{Z_t\} \sim$ IID, symmetric, $Z_1 \in D(\alpha)$, $0 < \alpha < 2$ and define

$$a_n = \inf\{x: P(|Z_1| > x) < n^{-1}\}$$

$$b_n = \inf\{x: P(|Z_1 Z_2| > x) < n^{-1}\}$$

Then,

$$\left(a_n^{-2} \sum_{t=1}^{n} Z_t^2,\ b_n^{-1} \sum_{t=1}^{n} Z_t Z_{t+1}, \ldots,\ b_n^{-1} \sum_{t=1}^{n} Z_t Z_{t+h}\right)$$

$$\xrightarrow{d} (S_0, S_1, \ldots, S_h),$$

where S_0 is stable($\alpha/2$) and $S_1, \ldots, S_h$ are stable(α).

Note: $S_0, S_1, \ldots, S_h$ are **independent** if $E|Z_1| = \infty$ and are **dependent** if $E|Z_1| < \infty$.

Special Case of Pareto Tails.

$$\left(n^{-2/\alpha} \sum_{t=1}^{n} Z_t^{\,2},\; (n \ln n)^{-1/\alpha} \sum_{t=1}^{n} Z_t\, Z_{t+1}, \ldots,\; (n \ln n)^{-1/\alpha} \sum_{t=1}^{n} Z_t\, Z_{t+h}\right)$$

$$\xrightarrow{d} (S_0, S_1, \ldots, S_h), \qquad (S_0, S_1, \ldots, S_h \text{ indep.})$$

Then,

$$\boxed{\begin{gathered}\left((n/\ln n)^{1/\alpha} \sum_{t=1}^{n} Z_t\, Z_{t+1}, \ldots,\; (n/\ln n)^{1/\alpha} \sum_{t=1}^{n} Z_t\, Z_{t+h}\right) \Big/ \sum_{t=1}^{n} Z_t^{\,2} \\ \xrightarrow{d} (S_1, \ldots, S_h)/\, S_0,\end{gathered}}$$

ARMA(p,q) Processes $\{X_t\}$.

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

$\{Z_t\} \sim \text{IID}(\alpha)$ with Pareto tails ($0 < \alpha < 2$)

Then,

$$(n / \ln n)^{1/\alpha} \left(\hat{\rho}(h) - \rho(h) \right) \xrightarrow{d} \sum_{j=1}^{\infty} \left(\rho(h+j) + \rho(h-j) - 2\rho(j)\rho(h) \right) S_j / S_0$$

where $\hat{\rho}(h)$ is the population ACF, and $\rho(h)$ is the sample ACF.

Implication: Estimation procedures which are inherently second order based will have scaling factor $(n / \ln n)^{1/\alpha}$. Examples are

- moment estimation (Davis & Resnick \`85, \`86)
- Yule-Walker estimation for AR's (Davis & Resnick \`85, \`86)
- Whittle estimate (and max Gaussian likelihood?) (Mikosch, Gadrich, Kluppelberg and Adler \`95)

M-Estimation

Example (AR(1)).

Data. $X_1, \ldots, X_n$

Model. $X_t = \phi_0 X_{t-1} + Z_t$,

$|\phi_0| < 1$, $\{Z_t\} \sim \mathrm{IID}(\alpha)$.

Loss function : $\rho(\ )$ with influence function $\psi(x) = \rho'(x)$ satisfying

1. Lipschitz of order $\tau > \max(\alpha - 1, 0)$.
2. $E|\psi(Z_1)| < \infty$ if $\alpha < 1$.
3. $E\,\psi(Z_1) = 0$ and $\mathrm{Var}(\psi(Z_1)) < \infty$ if $\alpha \geq 1$.

Minimize

$$T_n(\phi) = \sum_{t=1}^{n} \left(\rho(X_t - \phi\, X_{t-1}) - \rho(Z_t)\right)$$

$$= \sum_{t=1}^{n} \left(\rho(Z_t - (\phi - \phi_0)\, X_{t-1}) - \rho(Z_t)\right)$$

Set $v = n^{1/\alpha}(\phi - \phi_0)$, and put $S_n(v) = T_n(\phi_0 + v n^{-1/\alpha})$, i.e.,

$$S_n(v) = \sum_{t=1}^{n} \left(\rho(Z_t - v n^{-1/\alpha}\, X_{t-1}) - \rho(Z_t)\right)$$

Properties of $S_n(v) = \sum_{t=1}^{n} \left(\rho(Z_t - vn^{-1/\alpha} X_{t-1}) - \rho(Z_t)\right)$:

- $S_n(v)$ has convex sample paths.
- $S_n(v) \xrightarrow{d} S(v)$, (finite dimensional distribution convergence, where $S(v)$ is a random process).
- $S_n \xrightarrow{d} S$ on $C(R)$.

Result. (Davis, Knight and Liu `92)

$$\hat{v}_n := \operatorname{argmin}(S_n(v))$$

$$= n^{1/\alpha}(\hat{\phi} - \phi_0) \xrightarrow{d} \hat{v} := \operatorname{argmin}(S(v))$$

Example (MA(1)).

Data. $X_1, \ldots, X_n$

Model. $X_t = Z_t + \theta Z_{t-1}$,

$$|\theta_0| < 1,\ \{Z_t\} \sim \mathrm{IID}(\alpha)$$

LAD estimation:

Minimize

$$T_n(\theta) = \sum_{t=1}^{n} \left(|Z_t(\theta)| - |Z_t(\theta_0)|\right)$$

$$= \sum_{t=1}^{n} \left(|X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots (-\theta)^{t-1} X_1| - |Z_t(\theta_0)|\right)$$

Set $u = n^{1/\alpha}(\theta - \theta_0)$,

$$S_n(u) = T_n(\theta_0 + un^{-1/\alpha})$$

$$= \sum_{t=1}^{n} \left(|Z_t(\theta_0 + un^{-1/\alpha})| - |Z_t(\theta_0)|\right)$$

(Not a convex function of u!)

Linearize $Z_t(\theta_0 + un^{-1/\alpha})$ to get

$$S_n(u) \sim \sum_{t=1}^{n} \left(|Z_t(\theta_0) + un^{-1/\alpha} Z_t'(\theta_0)| - |Z_t(\theta_0)|\right)$$

where $-Z_t'(\theta_0)$ is the AR(1) process

$$Y_t = -\theta_0 Y_{t-1} + Z_t .$$

Result : Same limit result as in the AR(1) case, i.e.

$$n^{1/\alpha}(\hat{\theta}_{LAD} - \theta_0) \xrightarrow{d} \hat{u} = \text{argmin } S(u).$$

Linearized Version :

Initial estimate :

$$\hat{\theta}_0 = \theta_0 + O_p(n^{-1/2\alpha})$$

Objective Function:

$$T_n(\theta) = \sum_{t=1}^{n} \left(|Z_t(\hat{\theta}_0) + Z'_t(\hat{\theta}_0)(\theta - \hat{\theta}_0)| - |Z_t(\hat{\theta}_0)|\right).$$

Then

$$n^{1/\alpha}(\hat{\theta}_L - \theta_0) \xrightarrow{d} \hat{u} = \operatorname{argmin} S(u)$$

LAD-Estimation (Davis `95).

ARMA Model : $\phi(B)X_t = \theta(B)Z_t$, $\{Z_t\} \sim \text{IID}(\alpha)$.

Set $\boldsymbol{\beta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)^T$ and put

$$Z_t(\boldsymbol{\beta}) = \phi(B)X_t - \theta_1 Z_{t-1}(\boldsymbol{\beta}) - \cdots - \theta_q Z_{t-q}(\boldsymbol{\beta}),$$

where $Z_t(\boldsymbol{\beta}) = X_t := 0$ for $t < 1$.

Then with $\mathbf{v} = n^{1/\alpha}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$

(i) $S_n(\mathbf{v}) := \sum_{t=1}^{n} \left(|Z_t(\boldsymbol{\beta}_0 + \mathbf{v}n^{-1/\alpha})| - |Z_t(\boldsymbol{\beta}_0)| \right)$

$\xrightarrow{d} S(\mathbf{v})$, (in $C(\mathbf{R}^{p+q})$)

(ii) $\hat{\mathbf{v}}_{LAD} = n^{1/2}(\hat{\boldsymbol{\beta}}_{LAD} - \boldsymbol{\beta}_0) \xrightarrow{d} \hat{\mathbf{v}} = \text{argmin } S(\mathbf{v})$

Summary of Limit Results

Model:

$\phi(B)X_t = \theta(B)Z_t\ ,\ \ \{Z_t\} \sim \mathrm{IID}(\alpha),\ 0<\alpha < 2.$

LAD estimates:

$$n^{1/\alpha}(\hat{\beta}_{LAD} - \beta_0) \xrightarrow{d} \hat{\mathbf{v}}$$

Least squares estimates: $(n / \ln n)^{1/\alpha}(\hat{\beta}_{LS} - \beta_0) \xrightarrow{d} \mathbf{W}$ (limit same as the Whittle estimate).

Simulation results: Cauchy noise

	LS	LAD
AR(1) ϕ=.4	.395(.041)	.399(.015)
MA(1) θ=.8	.795(.049)	.794(.036)
ARMA(1,1) ϕ=.4	.399(.053)	.399(.026)
θ=.8	.781(.046)	.781(.033)

Bootstrapping the M-Estimate (Davis and Wu `95).

Data. $X_1, \ldots, X_n$

Model. $X_t = \phi_1 X_{t-1} + \qquad + \phi_p X_{t-p} + Z_t$, $\{Z_t\} \sim \mathrm{IID}(\alpha)$

M-estimate. $\hat{\phi}$

Estimated residuals. $\hat{Z}_t = X_t - \phi_1 X_{t-1} + \cdots - \phi_p X_{t-p}$

Bootstrap sample. $X_t^* = \hat{\phi}_1 X_{t-1}^* + \cdots + \hat{\phi}_p X_{t-p}^* + Z_t^*$

for $t = 1, \ldots, m_n$, where $\{Z_t^*\} \sim \mathrm{IID}(F_n)$, F_n = empirical df of $\hat{Z}_1, \ldots, \hat{Z}_n$.

BS M-estimate. $\hat{\phi}^*$

Result. If $m_n / n \longrightarrow 0$, then

$$P(m_n^{1/\alpha}(\hat{\phi}^* - \hat{\phi}) \in \bullet \mid \mathbf{X}_n) \xrightarrow{p} P(\hat{\mathbf{v}} \in \bullet) .$$

Removing the dependence on normalizing constants.

Let

$$M_n = \max\{|X_1|, \ldots, |X_n|\}$$

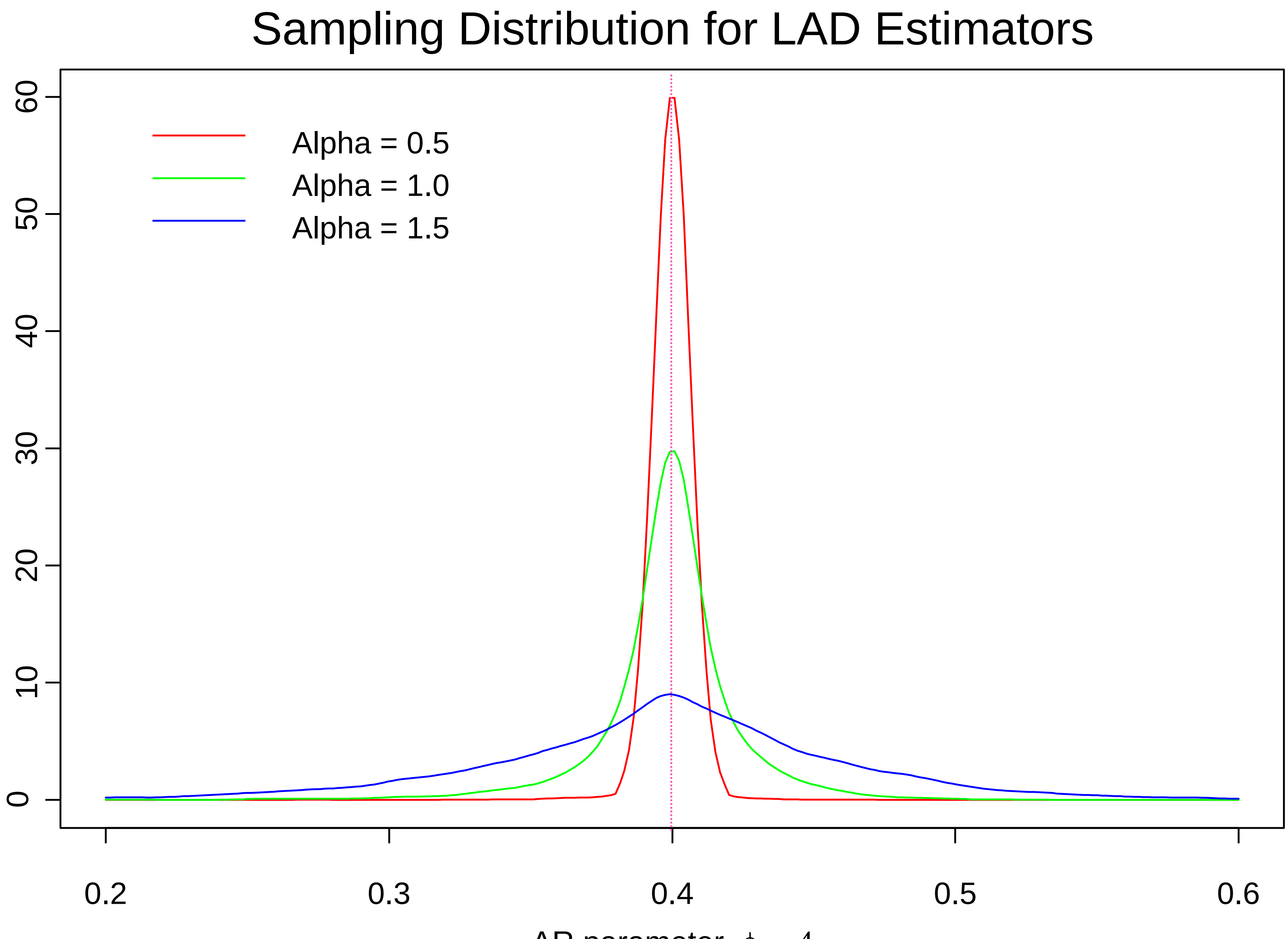
$$M_m^* = \max\{|X_1^*|, \ldots, |X_m^*|\}.$$

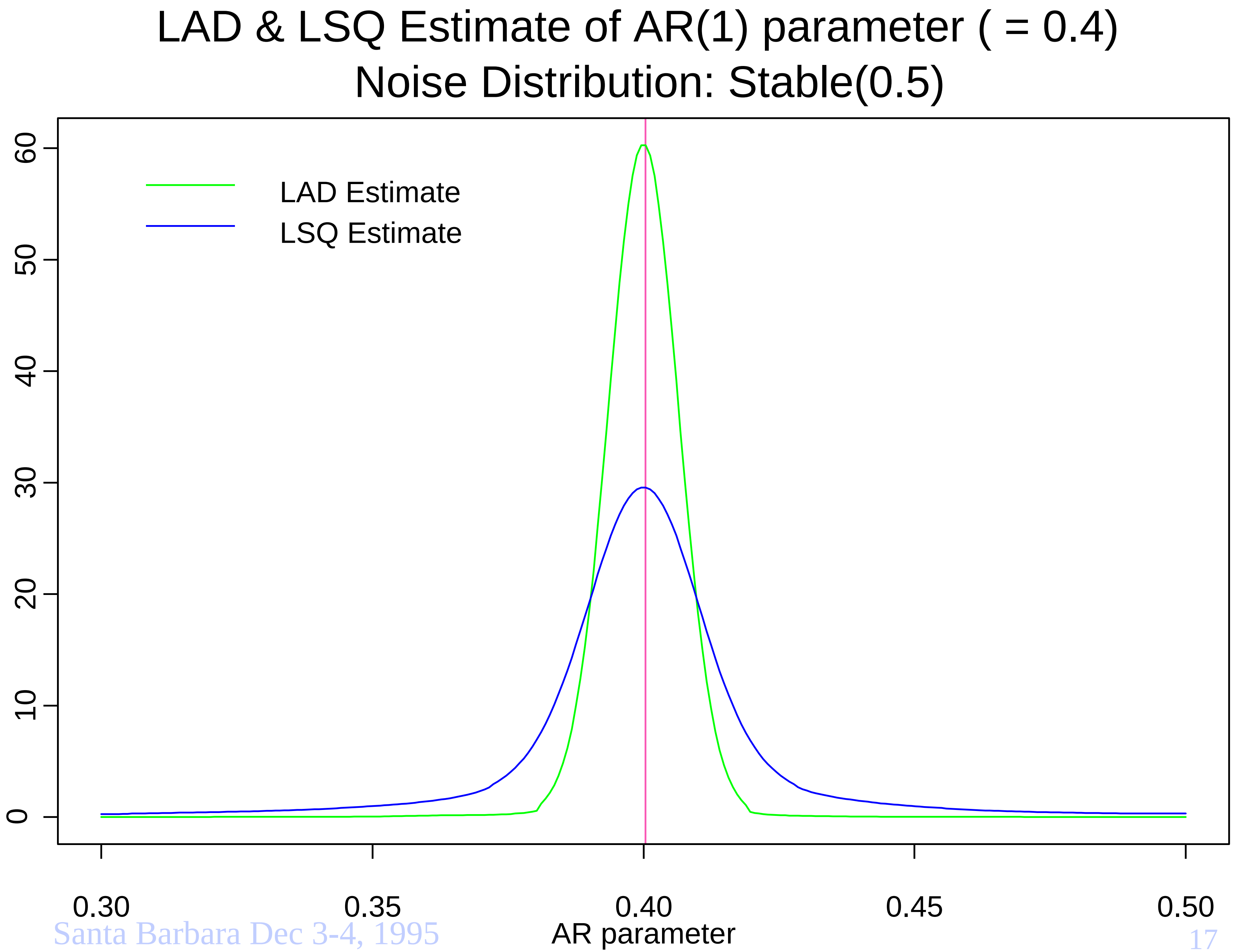
Then

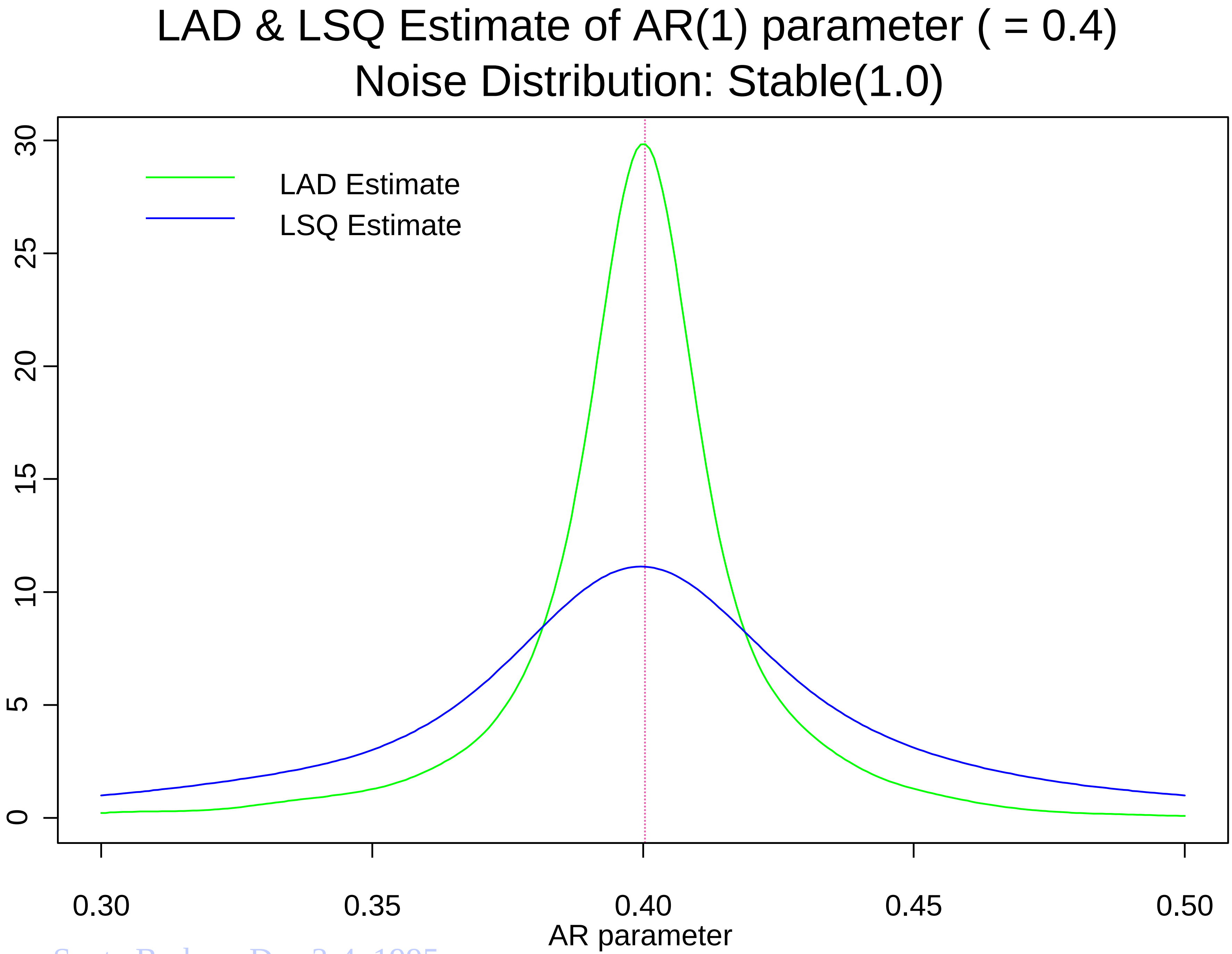
$$M_n(\hat{\phi} - \phi) \xrightarrow{d} \hat{\mathbf{w}}$$

and

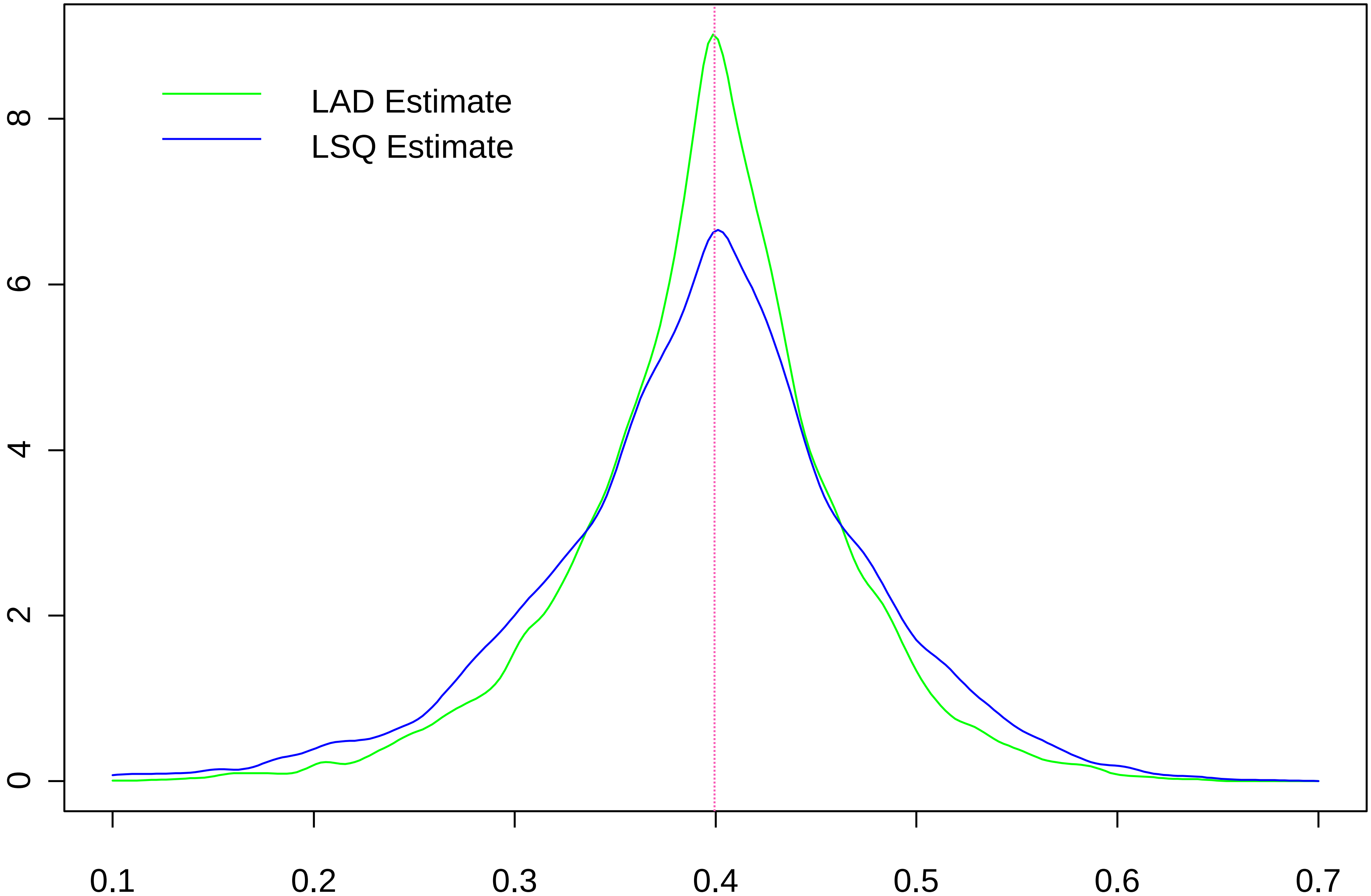
$$P(M_m^*(\hat{\phi}^* - \hat{\phi}) \in \bullet \mid \mathbf{X}_n) \xrightarrow{p} P(\hat{\mathbf{w}} \in \bullet) .$$

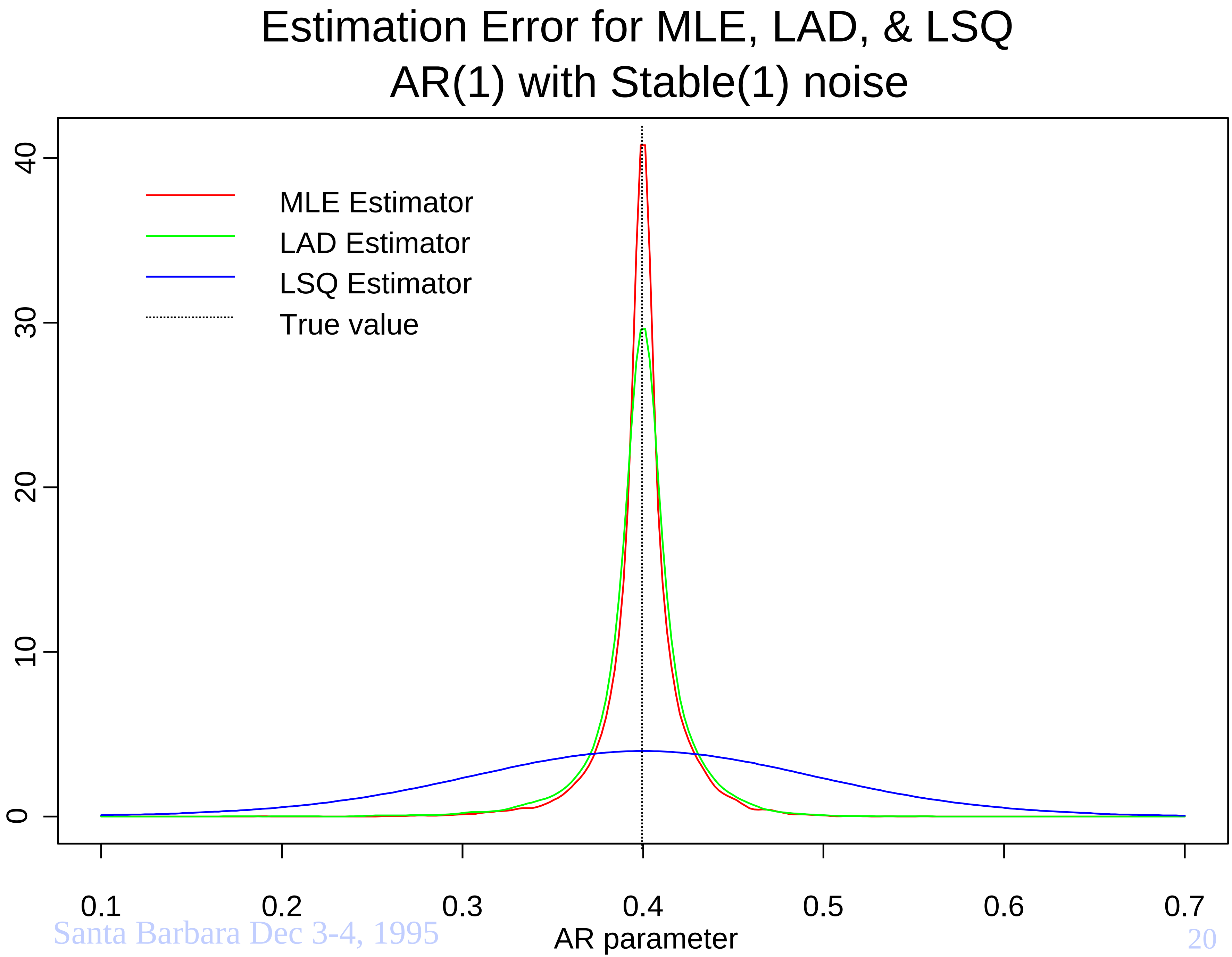

Sampling Distribution for LAD Estimators
Alpha = 0.5
Alpha = 1.0
Alpha = 1.5
0
10
20
30
40
50
60
0.2
0.3
0.4
0.5
0.6
AR parameter $\phi = .4$


LAD & LSQ Estimate of AR(1) parameter (= 0.4)
Noise Distribution: Stable(0.5)
LAD Estimate
LSQ Estimate
0
10
20
30
40
50
60
0.30
0.35
0.40
0.45
0.50
AR parameter


LAD & LSQ Estimate of AR(1) parameter (= 0.4)
Noise Distribution: Stable(1.0)
LAD Estimate
LSQ Estimate
0
5
10
15
20
25
30
0.30
0.35
0.40
0.45
0.50
AR parameter

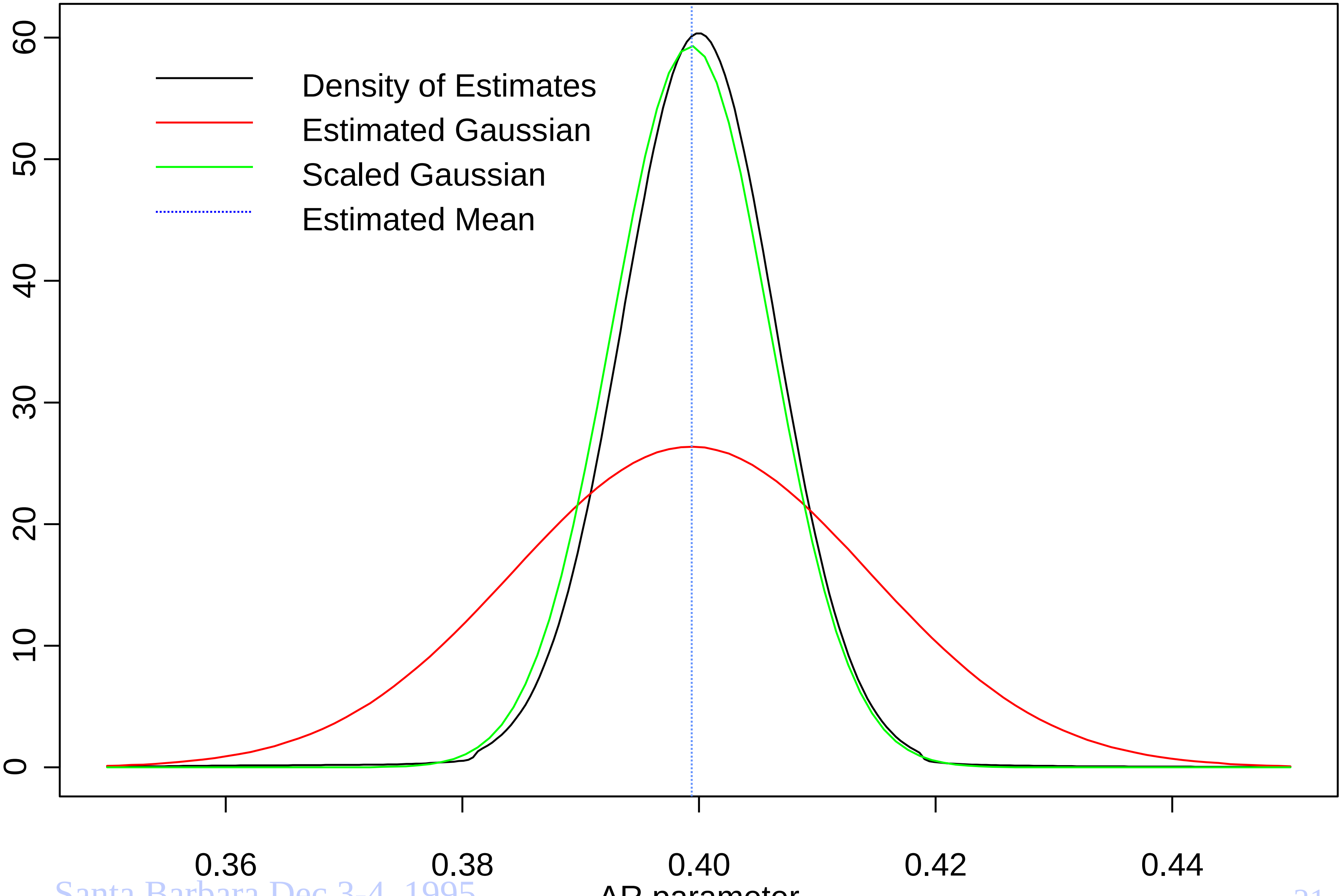
LAD & LSQ Estimate of AR(1) parameter (= 0.4)
Noise Distribution: Stable(1.5)



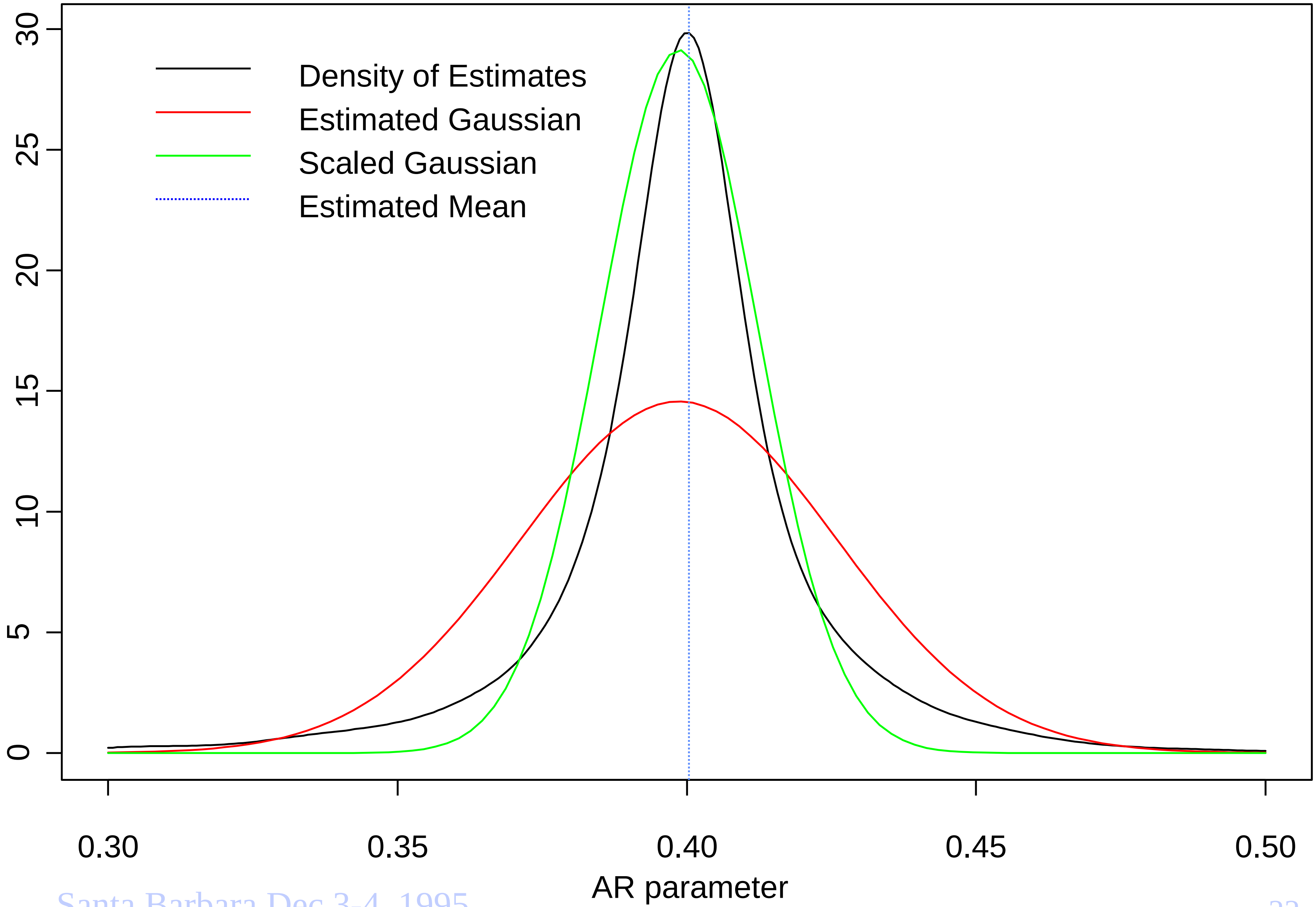


Estimation Error for MLE, LAD, & LSQ
AR(1) with Stable(1) noise
MLE Estimator
LAD Estimator
LSQ Estimator
True value
0
10
20
30
40
0.1
0.2
0.3
0.4
0.5
0.6
0.7
AR parameter

LAD Estimate of AR(1) parameter (= 0.4)
Noise Distribution: Stable(0.5)



LAD Estimate of AR(1) parameter (= 0.4)
Noise Distribution: Stable(1.0)



LAD Estimate of AR(1) parameter (= 0.4)
Noise Distribution: Stable(1.5)