Ziv-Zakai Bound for parameter estimation

Dinesh Ramasamy
dineshr@ece.ucsb.edu

Introduction

Consider the problem where we wish to estimate the scalar $\theta \in \mathbb{R}$ from noisy observations $\mathbf{y} = [y(n)] \in \mathbb{R}^N$. The model problem that we will use throughout is that of estimating the parameter θ from measurements corrupted by additive white Gaussian noise (AWGN):

$$y(n) = s(\theta, n) + z(n),\ n = 1, \ldots, N, \tag{1}$$

where $z(n)$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ and the prior on θ is $p(\theta)$. The *measurement model*, given by the conditional density $p(\mathbf{y}|\theta)$, is

$$\log p(\mathbf{y}|\theta) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{k=N}\left(y(k) - s(\theta, k)\right)^2. \tag{2}$$

Let $\hat{\theta}(\mathbf{y})$ be any *unbiased* estimator of θ. Denoting the vector $[s(\theta, k)]_{k=1}^{k=N}$ by $\mathbf{s}(\theta)$, $[\partial s(\theta, k)/\partial\theta]_{k=1}^{k=N}$ by $\partial\mathbf{s}(\theta)/\partial\theta$ and $[z(k)]_{k=1}^{k=N}$ by $\mathbf{z}$, the Cramér Rao Lower Bound (CRLB) is a lower bound on its variance and is given by the inverse of the Fisher information:

$$\begin{aligned}
\mathrm{FI}(\theta) &= \mathbb{E}_{\mathbf{y}|\theta}\left(\frac{\partial \log p(\mathbf{y}|\theta)}{\partial\theta}\left(\frac{\partial \log p(\mathbf{y}|\theta)}{\partial\theta}\right)^T\right) \\
&= \frac{1}{\sigma^4}\mathbb{E}_{\mathbf{y}|\theta}\left(\left(\left(\frac{\partial\mathbf{s}(\theta)}{\partial\theta}\right)^T(\mathbf{s}(\theta) - \mathbf{y})\right)\left(\left(\frac{\partial\mathbf{s}(\theta)}{\partial\theta}\right)^T(\mathbf{s}(\theta) - \mathbf{y})\right)^T\right) \\
&= \frac{1}{\sigma^4}\mathbb{E}_{\mathbf{y}|\theta}\left(\left(\frac{\partial\mathbf{s}(\theta)}{\partial\theta}\right)^T\mathbf{z}\mathbf{z}^T\ \frac{\partial\mathbf{s}(\theta)}{\partial\theta}\right) \\
&= \frac{1}{\sigma^2}\left\|\frac{\partial\mathbf{s}(\theta)}{\partial\theta}\right\|^2 \qquad \text{since } \mathbb{E}\ \mathbf{z}\mathbf{z}^T = \sigma^2\mathbb{I}_N
\end{aligned} \tag{3}$$

The CRLB, given thus by $\sigma^2\left\|\partial\mathbf{s}(\theta)/\partial\theta\right\|^{-2}$, gives a good prediction of estimation error in "low noise scenarios." We have the ML estimator for θ:

$$\hat{\theta}_{\mathrm{ML}}(\mathbf{y}) = \arg\max_{\phi}\log p(\mathbf{y}|\phi) = \arg\min_{\phi}\|\mathbf{y} - \mathbf{s}(\phi)\|^2 = \arg\min_{\phi}\|(\mathbf{s}(\theta) - \mathbf{s}(\phi)) + \mathbf{z}\|^2.$$

When the noise level is small, we expect that the minimizing ϕ will be in the neighborhood of the true value of θ (since $\|(\mathbf{s}(\theta) - \mathbf{s}(\phi)) + \mathbf{z}\| \approx \|\mathbf{s}(\theta) - \mathbf{s}(\phi)\|$). In such scenarios, the following argument applies: Suppose we have coarse estimate of the true value of the parameter θ given by θ_{coarse} (so that $|\theta - \theta_{\text{coarse}}|$ is small). One may linearize the problem in the following manner:

$$\begin{aligned}\mathbf{y} &= \mathbf{s}(\theta) + \mathbf{z}, \\ &= \mathbf{s}(\theta_{\text{coarse}}) + \frac{\partial \mathbf{s}(\theta_{\text{coarse}})}{\partial \theta}(\theta - \theta_{\text{coarse}}) + O\left((\theta - \theta_{\text{coarse}})^2\right) + \mathbf{z}.\end{aligned}$$

If $|\theta - \theta_{\text{coarse}}|$ is small enough so that the $O\left((\theta - \theta_{\text{coarse}})^2\right)$ term can be ignored, we can compute the MSE for this linear estimation problem:

$$\text{MSE}(\theta) = \sigma^2 \left\| \frac{\partial \mathbf{s}(\theta_{\text{coarse}})}{\partial \theta} \right\|^{-2}.$$

So, when we are given this extra information $\theta \approx \theta_{\text{coarse}}$, which we expect to be available via an ML estimate of θ in low noise scenarios, we can do as well as the CRLB. Are there situations when this bound is unreachable? Yes. The CRLB is optimistic in its predictions of estimation error in the "high noise scenario." In this lecture, we will go over one conceptually simple bound, the Ziv-Zakai Bound (ZZB) which gives a tight lower bound on estimation error in both low and high noise scenarios. One might ask, why bother with the CRLB? Why not use the ZZB always?

(i) Unlike the Cramér Rao Lower Bound, for which we only need the measurement model $p(\mathbf{y}|\theta)$, we also need the prior $p(\theta)$ on the parameter θ in order to evaluate the ZZB.

(ii) Computing the ZZB involves numerical integration, while the CRLB is typically easier to compute.

One of the main utilities of the ZZB is that it helps delineating the "high SNR regime" where the CRLB is an accurate prediction of achievable estimation error. We shall now go over the basic intuition behind the ZZB.

Ziv-Zakai Bound

Suppose we have a "good" estimator $\hat{\theta}(\mathbf{y})$ for θ in the model problem in (1), where θ can take any value in $\mathbb{R}$ with prior probability $p(\theta)$. This estimator can be used to construct a near-optimal detector for the following binary hypothesis testing problem:

Binary hypothesis testing problem: Suppose that we restrict θ to take two values $H_1 : \ \theta_1$ and $H_2 : \ \theta_2$ with prior probabilities $\Pr[H_1] = p(\theta_1)/(p(\theta_1) + p(\theta_2))$ and $\Pr[H_2] = 1 - \Pr[H_1]$, while retaining the measurement model in (1). The estimator $\hat{\theta}(\mathbf{y})$ does not know that θ is restricted to θ_1 and θ_2 and returns an estimate in $\mathbb{R}$. However, we may use a simple **nearest neighbor** detection strategy based on the estimate $\hat{\theta}(\mathbf{y})$, given by:

$$\left|\hat{\theta}(\mathbf{y}) - \theta_1\right| \underset{\theta_2}{\overset{\theta_1}{\lessgtr}} \left|\hat{\theta}(\mathbf{y}) - \theta_2\right|. \tag{4}$$

Let us denote the probability of error corresponding to this nearest-neighbor detector derived from the estimator $\hat{\theta}(\mathbf{y})$ (this error depends on the two values that θ can take, given by θ_1 and θ_2, and the estimator $\hat{\theta}(\mathbf{y})$) by $P_{\rm nn}^{\hat{\theta}}(\theta_1, \theta_2)$.

The optimal decision strategy for this binary hypothesis problem exploits the prior knowledge that θ can take only the two values θ_1 and θ_2. The optimal decision rule is the MAP rule given by:

$$\begin{aligned} \Pr[\theta_1|\mathbf{y}] &\underset{\theta_2}{\overset{\theta_1}{\gtrless}} \Pr[\theta_2|\mathbf{y}] \\ p(\mathbf{y}|\theta_1)p(\theta_1) &\underset{\theta_2}{\overset{\theta_1}{\gtrless}} p(\mathbf{y}|\theta_2)p(\theta_2) \\ \|\mathbf{y}-\mathbf{s}(\theta_1)\|^2 &\underset{\theta_2}{\overset{\theta_1}{\lessgtr}} \|\mathbf{y}-\mathbf{s}(\theta_2)\|^2 + 2\sigma^2 \log\left(\frac{p(\theta_1)}{p(\theta_2)}\right) \end{aligned} \tag{5}$$

We can compute the probability of error $P_{\rm opt}(\theta_1, \theta_2)$ for this *optimal* detection rule (5) for any pair of parameters θ_1 and θ_2 in terms of the CCDF of the standard normal random variable $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2)\mathrm{d}t$. This can be shown to be equal to:

$$\begin{aligned} P_{\rm opt}(\theta_1, \theta_2) &= \frac{p(\theta_1)}{p(\theta_1)+p(\theta_2)} Q\left(\frac{\|\mathbf{s}(\theta_1)-\mathbf{s}(\theta_2)\|}{2\sigma} + \frac{\sigma\log\left(p(\theta_1)/p(\theta_2)\right)}{\|\mathbf{s}(\theta_1)-\mathbf{s}(\theta_2)\|}\right) \\ &\quad + \frac{p(\theta_2)}{p(\theta_1)+p(\theta_2)} Q\left(\frac{\|\mathbf{s}(\theta_1)-\mathbf{s}(\theta_2)\|}{2\sigma} - \frac{\sigma\log\left(p(\theta_1)/p(\theta_2)\right)}{\|\mathbf{s}(\theta_1)-\mathbf{s}(\theta_2)\|}\right). \end{aligned} \tag{6}$$

Since the MAP rule (5) is *optimal*, $P_{\rm opt}(\theta_1, \theta_2)$ is the minimum probability of error that can be achieved for this detection problem. Therefore, however good an estimator $\hat{\theta}$ is, the nearest neighbor detector derived from it cannot beat the performance of the MAP rule:

$$P_{\rm nn}^{\hat{\theta}}(\theta_1, \theta_2) \geq P_{\rm opt}(\theta_1, \theta_2) \ \forall \theta_1, \theta_2 \ \& \text{ estimators } \hat{\theta}(\mathbf{y}). \tag{7}$$

Bounding estimation errors: One may wonder, how is all this related to *estimation* errors? Surprisingly, we can express the mean squared error (MSE) of the estimator $\hat{\theta}(\mathbf{y})$ in terms of the probability of detection error $P_{\rm nn}^{\hat{\theta}}(\theta_1, \theta_2)$ of the nearest neighbor detector derived from the estimator $\hat{\theta}(\mathbf{y})$! By replacing this probability of detection error $P_{\rm nn}^{\hat{\theta}}(\theta_1, \theta_2)$ with the probability of error of the MAP rule (optimal detection strategy) $P_{\rm opt}(\theta_1, \theta_2)$, we get a lower bound on the MSE which is independent of the estimator $\hat{\theta}(\mathbf{y})$ and this is the Ziv-Zakai Bound.

For any non-negative random variable X,

$$\mathbb{E}X^2 = \int_0^\infty x^2 p(x)\mathrm{d}x \underset{\text{by parts}}{\overset{\text{integration}}{=}} \underbrace{-x^2 \Pr[X \geq x]\Big|_{x=0}^{x=\infty}}_{0} + \int_0^\infty 2x \Pr[X \geq x]\mathrm{d}x,$$

and by a change of variables, this becomes

$$\mathbb{E}X^2 = \frac{1}{2}\int_0^\infty \Pr\left[X \geq \frac{x}{2}\right] x\mathrm{d}x.$$

This gives the following expression for the MSE $\mathbb{E}|\hat{\theta}(\mathbf{y}) - \theta|^2$ of the estimator $\hat{\theta}(\mathbf{y})$:

$$\text{MSE}\left(\hat{\theta}(\mathbf{y})\right) = \frac{1}{2}\int_0^\infty \Pr\left[\left|\hat{\theta}(\mathbf{y}) - \theta\right| \geq \frac{x}{2}\right] \; x\mathrm{d}x$$

and this is used to relate $\text{MSE}(\hat{\theta}(\mathbf{y}))$ to the probability of error $P_{\text{opt}}(\theta_1, \theta_2)$ of the optimal detection rule as follows:

$$\begin{aligned}
\Pr\left[\left|\hat{\theta}(\mathbf{y}) - \theta\right| \geq \frac{x}{2}\right] &= \Pr\left[\hat{\theta}(\mathbf{y}) - \theta \geq \frac{x}{2}\right] + \Pr\left[\hat{\theta}(\mathbf{y}) - \theta \leq -\frac{x}{2}\right] \\
&= \int_{-\infty}^{\infty} p(\varphi) \Pr\left[\hat{\theta}(\mathbf{y}) - \theta \geq \frac{x}{2}\middle| \theta = \varphi\right] \mathrm{d}\varphi \\
&\qquad + \int_{-\infty}^{\infty} p(\varphi) \Pr\left[\hat{\theta}(\mathbf{y}) - \theta \leq -\frac{x}{2}\middle| \theta = \varphi\right] \mathrm{d}\varphi \\
&= \int_{-\infty}^{\infty} p(\varphi) \Pr\left[\hat{\theta}(\mathbf{y}) - \theta \geq \frac{x}{2}\middle| \theta = \varphi\right] \mathrm{d}\varphi \\
&\qquad + \int_{-\infty}^{\infty} p(\varphi + x) \Pr\left[\hat{\theta}(\mathbf{y}) - \theta \leq -\frac{x}{2}\middle| \theta = \varphi + x\right] \mathrm{d}\varphi \\
&= \int_{-\infty}^{\infty} p(\varphi) \Pr\left[\hat{\theta}(\mathbf{y}) - \varphi \geq \frac{x}{2}\middle| \theta = \varphi\right] \mathrm{d}\varphi \\
&\qquad + \int_{-\infty}^{\infty} p(\varphi + x) \Pr\left[\hat{\theta}(\mathbf{y}) - \varphi \leq \frac{x}{2}\middle| \theta = \varphi + x\right] \mathrm{d}\varphi \\
&= \int_{-\infty}^{\infty} (p(\varphi) + p(\varphi + x)) \times \left\{ \frac{p(\varphi)}{p(\varphi) + p(\varphi + x)} \Pr\left[\hat{\theta}(\mathbf{y}) \geq \varphi + \frac{x}{2}\middle| \theta = \varphi\right] \right. \\
&\qquad\qquad \left. + \frac{p(\varphi + x)}{p(\varphi) + p(\varphi + x)} \Pr\left[\hat{\theta}(\mathbf{y}) \leq \varphi + \frac{x}{2}\middle| \theta = \varphi + x\right] \right\} \mathrm{d}\varphi
\end{aligned}$$

The term inside the curly braces, given by

$$\frac{p(\varphi)}{p(\varphi) + p(\varphi + x)} \Pr\left[\hat{\theta}(\mathbf{y}) \geq \varphi + \frac{x}{2}\middle| \theta = \varphi\right] + \frac{p(\varphi + x)}{p(\varphi) + p(\varphi + x)} \Pr\left[\hat{\theta}(\mathbf{y}) \leq \varphi + \frac{x}{2}\middle| \theta = \varphi + x\right],$$

is the probability of error of the nearest-neighbor detection strategy $P_{\text{nn}}^{\hat{\theta}}(\varphi, \varphi + x)$ when θ is restricted to φ and $\varphi + x$. To see this, notice that the following decision strategies are equivalent for $x \geq 0$:

$$S_1: \quad \text{Decide}(\mathbf{y}) = \begin{cases} \varphi & \hat{\theta}(\mathbf{y}) < \varphi + x/2 \\ \varphi + x & \text{otherwise} \end{cases}$$

$$S_2: \quad \text{Decide}(\mathbf{y}) = \begin{cases} \varphi & \left|\hat{\theta}(\mathbf{y}) - \varphi\right| < \left|\hat{\theta}(\mathbf{y}) - (\varphi + x)\right| \\ \varphi + x & \text{otherwise} \end{cases}$$

Therefore,

$$\Pr\left[\left|\hat{\theta}(\mathbf{y}) - \theta\right| \geq \frac{x}{2}\right] = \int_{-\infty}^{\infty} (p(\varphi) + p(\varphi + x)) \, P_{\text{nn}}^{\hat{\theta}}(\varphi, \varphi + x)\mathrm{d}\varphi. \tag{8}$$

Since $P_{\text{opt}}(\varphi, \varphi + x) \leq P_{\text{nn}}^{\hat{\theta}}(\varphi, \varphi + x)$ for all estimators $\hat{\theta}(\cdot)$, we have:

$$\Pr\left[\left|\hat{\theta}(\mathbf{y}) - \theta\right| \geq \frac{x}{2}\right] \geq \int_{-\infty}^{\infty} \left(p(\varphi) + p(\varphi + x)\right) P_{\text{opt}}(\varphi, \varphi + x) \mathrm{d}\varphi, \tag{9}$$

and using the above in the expression for MSE, we arrive at the ZZB:

$$\text{MSE}\left(\hat{\theta}(\mathbf{y})\right) \geq \frac{1}{2} \int_{x=0}^{x=\infty} \int_{\varphi=-\infty}^{\varphi=\infty} \left(p(\varphi) + p(\varphi + x)\right) P_{\text{opt}}(\varphi, \varphi + x) \mathrm{d}\varphi \; x \mathrm{d}x, \tag{10}$$

where $P_{\text{opt}}(\varphi, \varphi + x)$ is given by (6). Notice that the right hand side is independent of the estimator $\hat{\theta}(\mathbf{y})$ and depends only on the measurement model $p(\mathbf{y}|\theta)$ and the prior $p(\theta)$ on θ. This is thus a lower bound on the estimation error of *any* estimator $\hat{\theta}(\mathbf{y})$. It is worth reiterating that the ZZB (10) requires the prior $p(\theta)$ and is independent of the realization of θ, unlike the CRLB which is a function of θ, but does not depend on the prior $p(\theta)$.

The ZZB (10) can be tightened by noticing that the CCDF on the left hand side of (9) is non-increasing. Therefore, the bound on the CCDF given by the right hand side of (9) can also be "filled-in" to make it non-increasing. These ideas can also be extended to non-Gaussian and vector parameter estimation settings. We recommend [1] for details of these extensions.

Example

We illustrate these ideas using the problem of estimating the phase of a 3D helix (see Figure 1, left) in AWGN as an example. $s(\theta, 1) = \cos(\theta)$, $s(\theta, 2) = \sin(\theta)$, $s(\theta, 3) = \alpha\theta$ for some $\alpha > 0$ and $p(\theta) = 1/(2M\pi)$, if $\theta \in [0, 2M\pi)$ and zero otherwise ($M > 0$ is the number of rotations of the helix). The CRLB is given by $\sigma^2/(1 + \alpha^2)$ (plugging in the expression for $s(\theta, n)$ in (3)). To compute the ZZB we need the distance between $\mathbf{s}(\theta_1)$ and $\mathbf{s}(\theta_2)$,

$$\|\mathbf{s}(\theta_2) - \mathbf{s}(\theta_1)\|^2 = 4\sin^2\left((\theta_2 - \theta_1)/2\right) + \alpha^2(\theta_2 - \theta_1)^2.$$

This distance depends only on the difference between the two phases of the helix: $\theta_2 - \theta_1$. So, we define

$$D(x) = \sqrt{4\sin^2(x/2) + \alpha^2 x^2}.$$

Using $\|\mathbf{s}(\theta_1) - \mathbf{s}(\theta_2)\| = D(\theta_2 - \theta_1)$ and the prior on θ in the expression for $P_{\text{opt}}(\theta_1, \theta_2)$, we have:

$$P_{\text{opt}}(\theta_1, \theta_2) = \begin{cases} Q\left(\frac{D(\theta_2 - \theta_1)}{2\sigma}\right) & \theta_1, \theta_2 \in [0, 2M\pi] \\ 0 & \text{otherwise} \end{cases}.$$

To compute the ZZB (10), we need $P_{\text{opt}}(\varphi, \varphi + x)$ for $x > 0$ and this is given by

$$\begin{aligned} P_{\text{opt}}(\varphi, \varphi + x) &= \begin{cases} Q\left(\frac{D(x)}{2\sigma}\right) & \varphi, (\varphi + x) \in [0, 2M\pi] \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} Q\left(\frac{D(x)}{2\sigma}\right) & \varphi \in [0, 2M\pi - x], x \leq 2M\pi \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

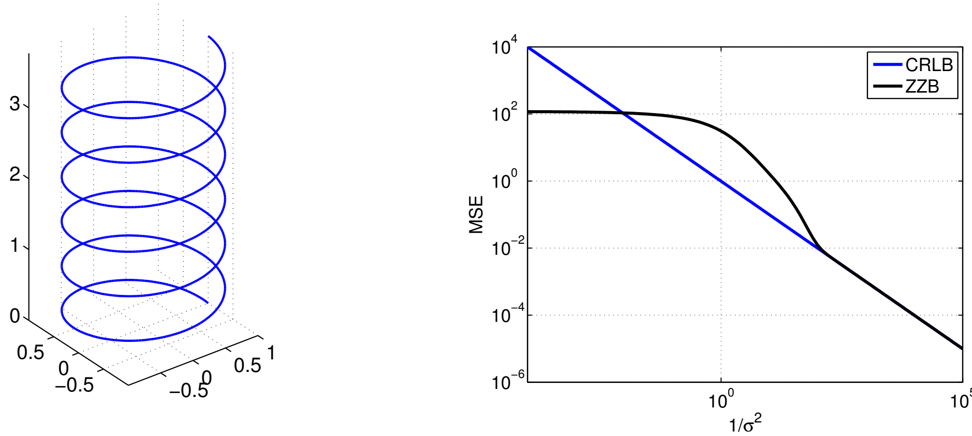

Figure 1: Example problem: Estimation the phase $\theta \in [0, 12\pi)$ of a 3D helix with $M = 6$ rotations and $\alpha = 0.1$ (drawn on the left). The prior on θ is uniform. The bounds on the MSE are plotted on the right as function of $1/\sigma^2$

Plugging this in the ZZB (10) and remembering that $p(\theta) = 1/(2M\pi)$ (for $\theta \in [0, 2M\pi)$),

$$\begin{aligned} \text{MSE}\left(\hat{\theta}(\mathbf{y})\right) &\geq \frac{1}{2}\int_{x=0}^{x=\infty}\int_{\varphi=-\infty}^{\varphi=\infty}\left(p(\varphi)+p(\varphi+x)\right)P_{\text{opt}}(\varphi,\varphi+x)\mathrm{d}\varphi\ x\mathrm{d}x \\ &= \frac{1}{2M\pi}\int_{x=0}^{x=2M\pi}\int_{\varphi=0}^{\varphi=2M\pi-x} Q\left(\frac{D(x)}{2\sigma}\right)\mathrm{d}\varphi\ x\mathrm{d}x \\ &= \int_{x=0}^{x=2M\pi}\left(1-\frac{x}{2M\pi}\right)Q\left(\frac{D(x)}{2\sigma}\right)\ x\mathrm{d}x \end{aligned}$$

We evaluate the CRLB and the ZZB for $\alpha = 0.1$, $M = 6$ and plot it as a function of the SNR $1/\sigma^2$ in Figure 1. Notice that for a range of SNRs, the ZZB is orders of magnitude tighter than the CRLB.

Exercise

Play around with the code below by changing `alpha`. How do the two figures change qualitatively?

Computing these bounds

```
%% EXPRESSIONS FOR THE CRLB AND ZZB
Distance = @(x, alpha) sqrt(4*(sin(x/2)).^2 + alpha^2*x.^2);
ZZB_fun = @(sigma, alpha, M) 0.5*integral(@(x) x.*(1-x/(2*M*pi)).*...
    erfc(Distance(x,alpha)/(2*sqrt(2)*sigma)), 0, 2*M*pi, 'AbsTol', 1e
        -32); %NUMERICAL INTEGRATION
CRLB_fun = @(sigma, alpha) sigma.^2/(1+alpha^2);

%% SET PARAMETERS
alpha = 0.1;
M = 6;

%% PLOT THE RESPONSE
vec_s = @(theta,alpha) [cos(theta) sin(theta) alpha*theta];
theta_sample = linspace(0, 2*pi*M, 1000);
vec_s_sample = vec_s(theta_sample(:),alpha);
figure; plot3(vec_s_sample(:,1), vec_s_sample(:,2), vec_s_sample(:,3));
axis equal;

%% COMPUTE THE CRLB AND ZZB
sigma = fliplr(logspace(-2.5, 2, 100));
CRLB = CRLB_fun(sigma, alpha);
ZZB = zeros(size(sigma));
for count = 1:length(sigma)
    ZZB(count) = ZZB_fun(sigma(count), alpha, M);
end

%% PLOT CRLB AND ZZB
figure; loglog(1./(sigma.^2), CRLB, 'b-'); hold on;
loglog(1./(sigma.^2), ZZB, 'k-');
xlabel('1/\sigma^2');
ylabel('MSE');
legend({'CRLB', 'ZZB'});
```

References

[1] K. Bell, Y. Steinberg, Y. Ephraim, and H. Van Trees. Extended Ziv-Zakai lower bound for vector parameter estimation. *Information Theory, IEEE Transactions on*, 43(2):624–637, Mar 1997.