Exercise on MaxLik problems

By Zdeněk Hradil

1. Radon and inverse Radon transformation

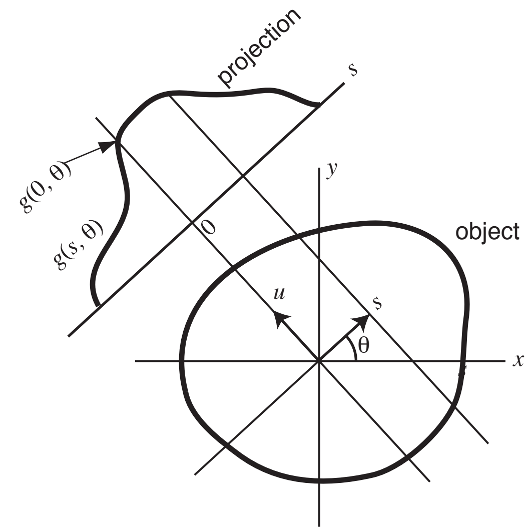
- Derive the analytical expression for Radon and Inverse Radon Transformations.

Geometry: Projection along the line

$$x \cos\theta + y \sin\theta - s = 0$$

Radon transformation

$$g(s,\theta) = \iint_{-\infty}^{\infty} f(x,y)\, \delta(x\cos\theta + y\sin\theta - s)\,dx\,dy$$



For obtaining Radon transformation as a Ray sum, define rotated system of coordinates

$$\begin{pmatrix} s \\ u \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Hence

$$x = s\cos\theta - u\sin\theta$$
$$y = s\sin\theta + u\cos\theta$$

$$\int \delta(s)ds = 1$$

Radon transformation as a Ray sum

$$g(s,\theta) = \int_{-\infty}^{\infty} f(s\cos\theta - u\sin\theta,\, s\sin\theta + u\cos\theta)\, du$$

Inverse Radon transformation follows directly from the definition

$$G_\theta(\xi) = \int_{-\infty}^{\infty} g(s,\theta)\exp(-i2\pi\xi s)ds$$

$$G_\theta(\xi) = F(\xi\cos\theta, \xi\sin\theta)$$

$$f(x,y) = \int_0^\infty \int_0^{2\pi} \xi d\xi d\theta F(\xi\cos\theta, \xi\sin\theta) e^{2\pi i(x\xi\cos\theta + y\xi\sin\theta)}$$
$$= \int_0^\infty \int_0^{2\pi} \xi d\xi d\theta G_\theta(\xi) e^{2\pi i(x\xi\cos\theta + y\xi\sin\theta)}$$

2. Statistical interpretation of measured results

- Assume the measurement of quadrature operator in some (unknown) coherent quantum state with real amplitude. The measurement was done 3times and the values x_1,x_2, and x_3 have been detected. Could the signal come from vacuum state? What could be the best guess??
- Hint: Assuming the normalization $x = (1/2)^{1/2}(a + a^{+})$, the probability density reads

$$p(x) = \frac{1}{\sqrt{\pi}} \exp[-(x - \sqrt{2}\alpha)^2]$$

- Solution: The likelihood for the measured state being in coherent state is

Hence the ratio is

$$\mathcal{L} = \prod_i \exp[-(x_i - \sqrt{2}\alpha)^2]$$

$$\mathcal{L}/\mathcal{L}(0) = \exp[2\sqrt{2}|\alpha| \sum_i x_i - 2n|\alpha^2|]$$

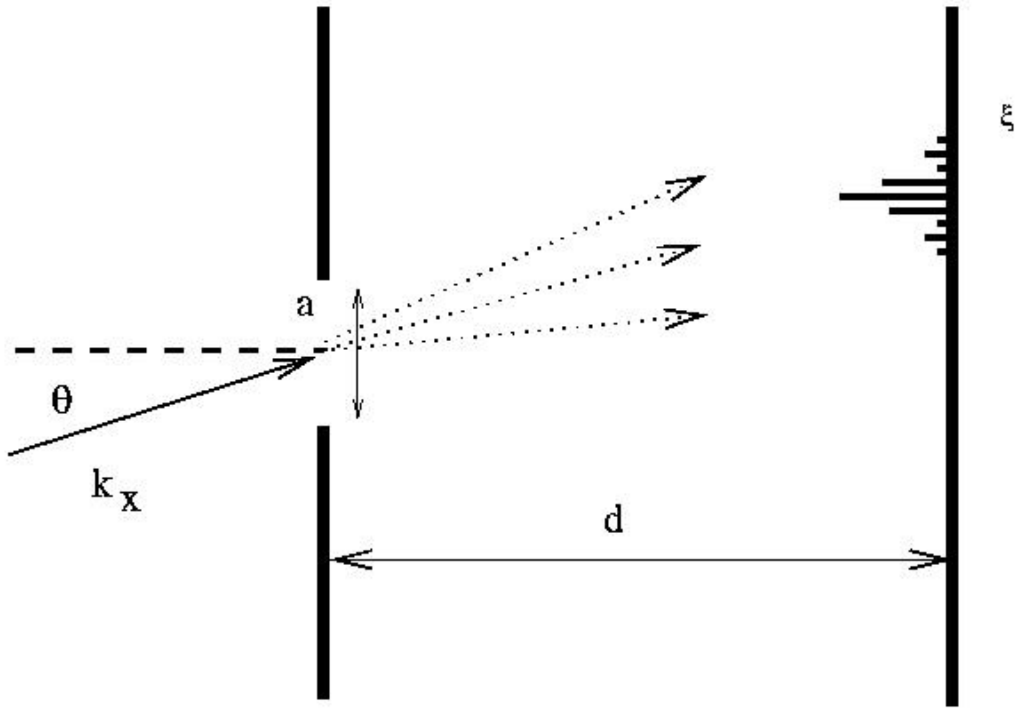
$$= \exp[-2n[|\alpha| - \frac{\sum_i x_i}{\sqrt{2n}}]^2 \exp((\sum_i x_i)^2/n)$$

Moral: Yes, indeed, data might come almost from any state but sometimes it is not too much likely that it really happened. For particular values x = 1,2,3 the optimal state is about 106 times more likely than the vacuum state.

3. Fisher information for the diffraction on the slit

- Derive the uncertainty principle and Fisher information from the model of 1D diffraction on the slit. Discus the “difference” between the measurement and estimation

Motivation: Diffraction on the slit



Detection on the screen may be used as geometrical estimate for impulse since $\theta = \xi/d$ and $p_x = h \sin\theta/\lambda$

Diffraction continues 1 …

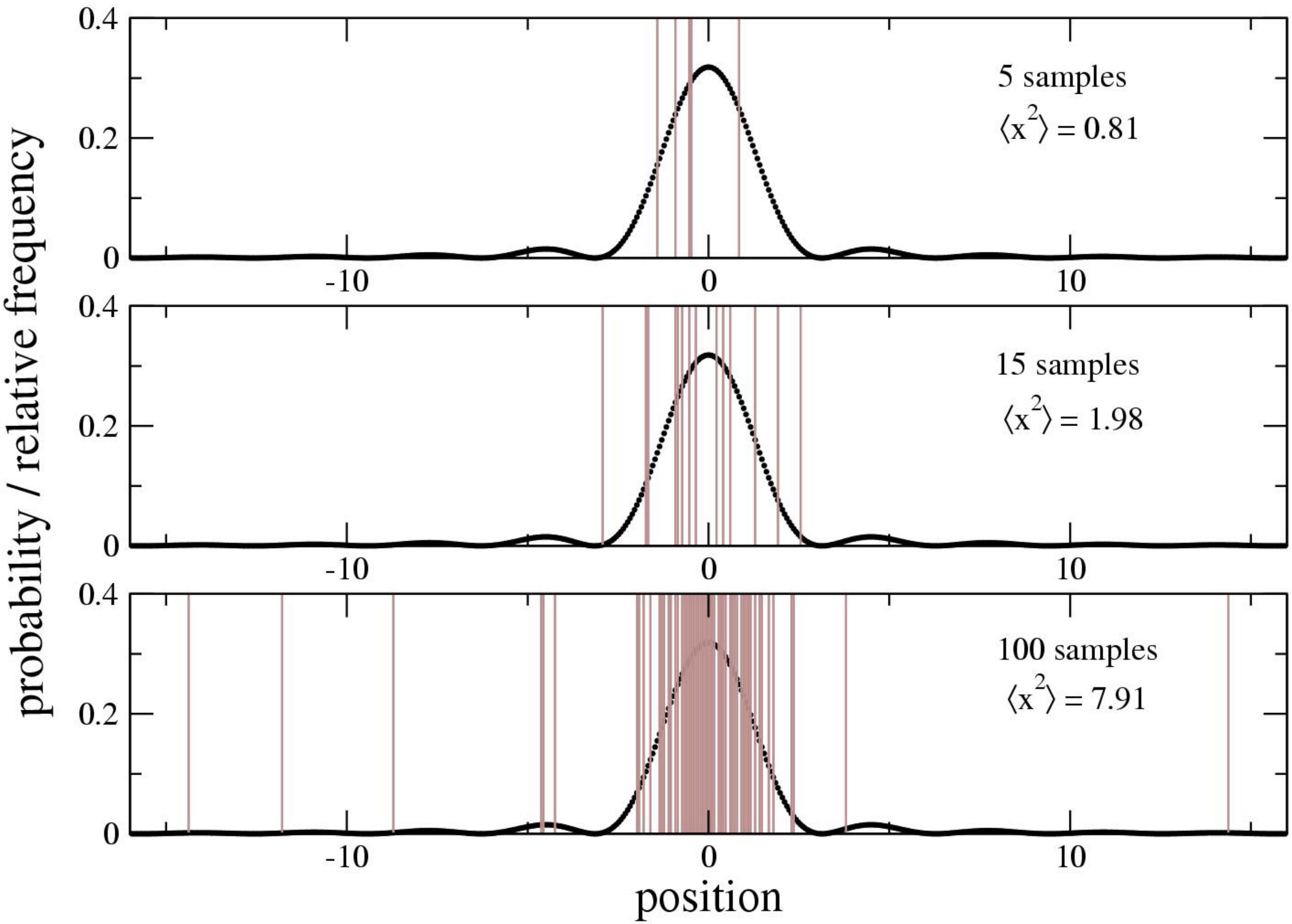
- **The uncertainty is given by wave theory**

  $P(\mu|\nu) = \pi^{-1} \, \mathrm{sinc}^2(\mu - \nu); \quad \mu = \xi \, (\pi a/\lambda d), \; \nu = p_x a/2\hbar$

- **Straightforward but wrong argumentation based on the first minimum of sinc function gives**

$\Delta x = a/2 \, , \, \Delta p_x = (h/a)$ and therefore $\Delta x \, \Delta p_x \sim h/2$!

- But the correctly calculated variance of sinc^2 function gives the infinite width !!
- The estimate of p_x based on single event will be very uncertain !!!
- The remedy is to accumulate the events and relate the estimate to some collective variable (=centre of mass of the interference pattern)
- Proper estimation theory should be formulated with the mathematical statistics.



probability / relative frequency
position
5 samples
⟨x^2⟩ = 0.81
15 samples
⟨x^2⟩ = 1.98
100 samples
⟨x^2⟩ = 7.91
0.4
0.2
0
-10
0
10

Diffraction continues 2 …

- The prediction should be based on some posterior distribution

$$P(v)_{post} = \Pi_\mu \; p(\mu|v)^{N_\mu} = \exp[\Sigma_\mu N_\mu \log p(\mu|v) \;].$$

Here v is our estimate of some true value $v_{true,}$ which is hidden in detected data μ

- Note: product of detected probabilities is denoted as likelihood L and its logarithm in exponential is called log-likelihood log L
- Significant sampling (N large) $\quad N_\mu = N\, p(\mu|v_{true})$
- Gaussian approximation of log L as the expansion near v_{true} :

$\Sigma_\mu N_\mu \log p(\mu|\, v) \sim \; N \,\Sigma_\mu\, p(\mu|v_{true}) \log p(\mu|v) \sim$

(1st term) $\quad N\, \Sigma_\mu\, p(\mu|v_{true}) \log p(\mu|\, v_{true})$

(2nd term) $\quad + N\, \Sigma_\mu\; p(\mu|v_{true}) \;\; \partial_v \log p(\mu|v)|_{true} \;\; (v-v_{true})$

(3rd term) $\quad + \frac{1}{2}\, N\, \Sigma_\mu\, p(\mu|v_{true})\; \partial^2{}_v \log p(\mu|v)|_{true} \;\; (v-v_{true})^2$

Diffraction continues 3 ...

1st term is entropy $S = \Sigma_\mu p(\mu|v_{true}) \log p(\mu| v_{true})$

2nd term is zero since $\Sigma_\mu \; p(\mu|v_{true}) \; \partial_v \log p(\mu|v)|_{true} =$
$\Sigma_\mu \partial_v p(\mu|v)|_{true} \; (v-v_{true}) = (v-v_{true}) \; \partial_v 1 = 0$

3rd term similarly gives the only nonzero contribution

$$F = N \Sigma_\mu p(\mu|v_{true}) [\partial_v \log p(\mu|v)|_{true}]^2$$
$$= N \Sigma_\mu p(\mu|v_{true})^{-1} [\partial_v p(\mu|v)|_{true}]^2$$

F = Fisher information

$$L \sim \exp(S) \exp[- \tfrac{1}{2} F (v-v_{true})^2]$$

This means that parameter estimation is done with the precision 1/F !

Diffraction continues 4 …

Believe or not Fisher information is remedy for uncertainty relations on the slit!

$(\Delta x)^2 = a^2/12$

$(\Delta v)^2 = (a/2\hbar)^2 (\Delta p_x)^2$ and $F = 4\pi^{-1} \int d\mu\, [\partial_\mu \operatorname{sinc} \mu]^2 = 4/3$

and therefore $\Delta x\, \Delta p_x = \hbar/2$!

This is not an accident but a consequence of Cramer-Rao inequalities (N=1):

Unbiased estimator : $\Sigma_\mu\, p(\mu|v_{true})\,(v - v_{true}) = 0 \quad / \partial v_{true}$

$$\Sigma_\mu\, \partial v_{true}\, p(\mu|v_{true})\,(v - v_{true}) = 1 \quad \text{/Cauchy-Schwarz inequality}$$

$$\Sigma_\mu\, [p(\mu|v_{true})]^{-1/2}\, \partial v_{true}\, p(\mu|v_{true})\, [p(\mu|v_{true})]^{1/2}\,(v - v_{true}) = 1$$

$$(\Delta v)^2\, F \geq 1$$

Some pedagogical remarks …

$$\Delta A \, \Delta B \geq \tfrac{1}{2} \, |[A,B]|$$

- The meaning of Heisenberg uncertainty principle is pedagogically confusing. Does it mean the constraints on measurement? Which one? Both?
- No, this is the constraint on possible quantum states (see the derivation or see the condition for covariance matrix).
- Heisenberg uncertainty is weaker than Cramer-Rao inequality

$$(\Delta v)^2 \; F \geq 1$$

- Cramer-Rao can be formulated even for simultaneous estimation (measurement) of several parameters.

4. **MaxLik solution**

- Derive the condition for MaxLik extremal state.
- Hint: Use the inequality between geometric and arithmetic means

Log-likelihood for generic measurement $p_i = Tr(\rho A_i)$

$$L(\rho) = \Pi_i \, p_j^{Ni}$$

Normalization $Tr(\rho) = 1$

Constraint $\rho \geq 0$

Maximize the likelihood !!!

Jensen inequality (inequality between geometric and arithmetic means) $\Pi_i \, (x_i/a_i)^{fi} \leq \sum_i f_i \, x_i/a_i$

$$L(\rho)^{1/N} = \Pi_i \, p_j^{fi} \leq (\Pi_i \, a_i^{fi}) \, Tr(R \, \rho)$$

$$R = \sum_i (f_i/a_i) \, A_i$$

Let us chose for extreme $a_i = Tr(\rho A_i)$

Extremal equation $R \, \rho = \rho$

Easy derivation

Differentiate formally the Log-likelihood with the constraint

$$\log L(\rho) = \sum_i N_i \log p_j(\rho) - \lambda\, Tr(\rho) \qquad /\partial\rho_{kl}$$

$$\sum_i N_i/p_j(\rho) \quad (A_i)_{kl} \quad |k\rangle\langle l| - \lambda\, \delta_{kl}\, |k\rangle\langle l| = 0 \qquad /\rho$$

$$\sum_i N_i/p_j(\rho)\, A_i\, \rho = \lambda\, \rho \qquad /Tr\rho = 1$$

$$R\, \rho = \rho$$

Other hints:

$\rho = \sum_i \lambda_i\, |\varphi_i\rangle\langle\varphi_i|$, $\partial\langle\varphi_i|\ \left[\, \langle\varphi_i| A_j\, |\varphi_i\rangle \right] = A_j\, |\varphi_i\rangle$;

$\rho = \Omega\Omega^\dagger \qquad \partial\, \Omega^\dagger\ Tr(A_j\, \Omega\Omega^\dagger\,) = A_j\, \Omega$

(Log)-likelihood is convex functional over the convex manifold of density matrices = convex optimization

5. Normalization of the likelihood

- Convince yourself that likelihood should be properly normalized. Conclude the consequences if it is not the case. Is the measurement in James et. al. , PRA 64, 052312 (2001) properly normalized? Calculate the corresponding *G* operator. How would you correct the paper James et al., *Measurement of qubits*, PRA 64, 052312 (2001)?
- Hint: Assume the measurement of single qbit corresponding just to single projections along +x, +y and +z axis of Stern-Gerlach apparatus. Such measurement is not complete nor normalized to 1 (show !). If non-normalized likelihood is used for a generic mixed state, then MaxLik estimate always tend to be a pure state …

ν	Mode 1	Mode 2	h_1	q_1	h_2	q_2
1	$\|H\rangle$	$\|H\rangle$	45^o	0	45^o	0
2	$\|H\rangle$	$\|V\rangle$	45^o	0	0	0
3	$\|V\rangle$	$\|V\rangle$	0	0	0	0
4	$\|V\rangle$	$\|H\rangle$	0	0	45^o	0
5	$\|R\rangle$	$\|H\rangle$	22.5^o	0	45^o	0
6	$\|R\rangle$	$\|V\rangle$	22.5^o	0	0	0
7	$\|D\rangle$	$\|V\rangle$	22.5^o	45^o	0	0
8	$\|D\rangle$	$\|H\rangle$	22.5^o	45^o	45^o	0
9	$\|D\rangle$	$\|R\rangle$	22.5^o	45^o	22.5^o	0
10	$\|D\rangle$	$\|D\rangle$	22.5^o	45^o	22.5^o	45^o
11	$\|R\rangle$	$\|D\rangle$	22.5^o	0	22.5^o	45^o
12	$\|H\rangle$	$\|D\rangle$	45^o	0	22.5^o	45^o
13	$\|V\rangle$	$\|D\rangle$	0	0	22.5^o	45^o
14	$\|V\rangle$	$\|L\rangle$	0	0	22.5^o	90^o
15	$\|H\rangle$	$\|L\rangle$	45^o	0	22.5^o	90^o
16	$\|R\rangle$	$\|L\rangle$	22.5^o	0	22.5^o	90^o

Table 1

TABLE 1: The tomographic analysis states used in our experiments. The number of coincidence counts measured in projections measurements provide a set of 16 data that allow the density matrix of the state of the two modes to be estimated. We have used the notation $|\mathrm{D}\rangle \equiv (|\mathrm{H}\rangle + |\mathrm{V}\rangle)/\sqrt{2}$, $|\mathrm{L}\rangle \equiv (|\mathrm{H}\rangle + i|\mathrm{V}\rangle)/\sqrt{2}$ and $|\mathrm{R}\rangle \equiv (|\mathrm{H}\rangle - i|\mathrm{V}\rangle)/\sqrt{2}$. Note that, when the measurement are taken in the order given by the table, only one waveplate angle had to be changed between each measurement.

6. Resource analysis for tomography, quantum computing and diagnostics with 5 q-bits

- To control the quantum system means to control all relevant errors….

- Pure state in dimension d: 2d -1 real parameters

Estimation is not a convex problem…

- Density matrix $d^2 - 1$ real parameters

Fisher info matrix: $\frac{1}{2}(d^2-1)(d^2-2)$ real parameters

- CP maps: $d^2(d^2-1)$ real parameters

Fisher info matrix for CP maps: $\frac{1}{2}d^2(d^2-1)(d^4 - d^2 - 1)$ real parameters

Quantum computation with 5 qbits: $d = 2^5 = 32$

Quantum state: ~ 10^3 parameters

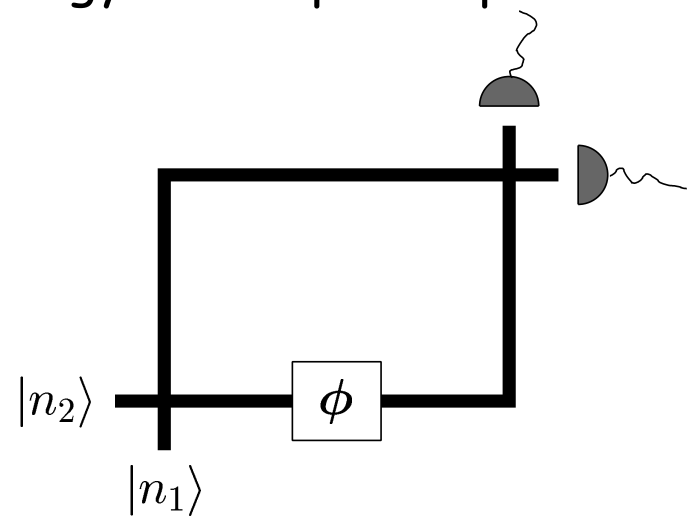
Fisher info: ~ 10^6 parameters

CP maps: ~ 10^6 parameters

Fisher info of CP maps: ~ 10^{12} parameters

7. Fisher information in quantum interferometry

- Considering the model of Holland, Burnett with the interferometer triggered by two n-Fock states, investigate the resolution limit and discuss the strategy when optimal phase resolution can be achieved.



Detected signal difference: q
($n_1 = n_2 = r$)

Detected statistics
(associated Legendre polynomials):

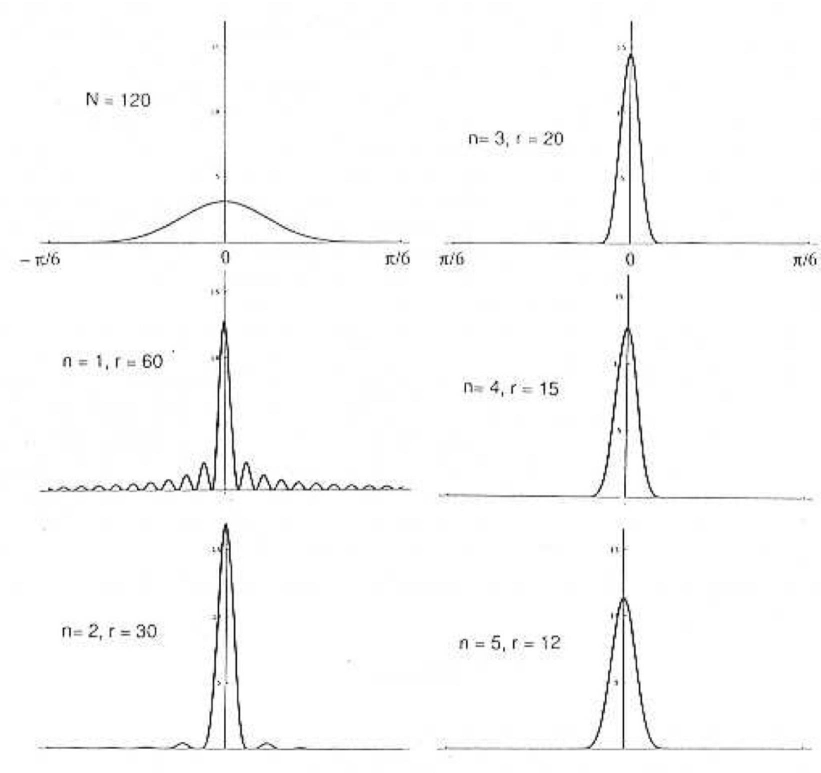
$$P(2q|\theta) = \frac{(r-q)!}{(r+q)!}[P_r^q(\cos\theta)]^2$$

Approximation for q=0

$$P(0|\theta) \approx [J_0(r\theta)]^2$$

Phase estimation after "single" detection is not a good idea since

$$(\Delta\phi)^2 \approx (\frac{1}{r})^2 \frac{\int_0^{r\pi/2} dx x^2 J_0^2(x)}{\int_0^{r\pi/2} dx J_0^2(x)} \propto \frac{1}{\log r}$$



Optimum: repeat n = 4

$$(\Delta\phi)^2 \approx \frac{4n}{N^2}$$