BME I5000: Biomedical Imaging

Lecture 7
Maximum Likelihood Reconstruction

Lucas C. Parra, parra@ccny.cuny.edu

Blackboard: http://cityonline.ccny.cuny.edu/

Schedule

1. Introduction, Spatial Resolution, Intensity Resolution, Noise
2. X-Ray Imaging, Mammography, Angiography, Fluoroscopy
3. Intensity manipulations: Contrast Enhancement, Histogram Equalisation
4. Computed Tomography
5. Image Reconstruction, Radon & Fourier Transform, Filtered Back Projection
6. Nuclear Imaging, PET and SPECT
7. Maximum Likelihood Reconstruction
8. Magnetic Resonance Imaging
9. Fourier reconstruction, k-space, frequency and phase encoding
10. Optical imaging, Fluorescence, Microscopy, Confocal Imaging
11. Enhancement: Point Spread Function, Filtering, Sharpening, Wiener filter
12. Segmentation: Thresholding, Matched filter, Morphological operations
13. Pattern Recognition: Feature extraction, PCA, Wavelets
14. Pattern Recognition: Bayesian Inference, Linear classification

Random Variables - Moments

review

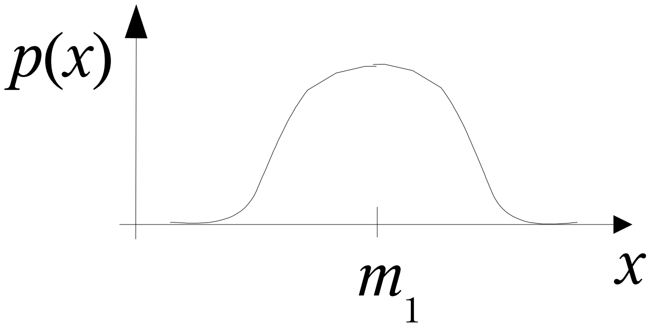
Expected value $E[f(X)]$ or **ensemble average** is defined as

$$E[f(X)]=\int_{-\infty}^{\infty} dx\, p(x) f(x)$$

Moment m_n of order n is the expected value

$$m_n=E[X^n]=\int_{-\infty}^{\infty} dx\, p(x) x^n$$

First moment is the **mean**



$$m_1=E[X]=\int_{-\infty}^{\infty} dx\, p(x) x$$

Second moment is the **power**

$$m_2=E[X^2]$$

Random Variables - Moments

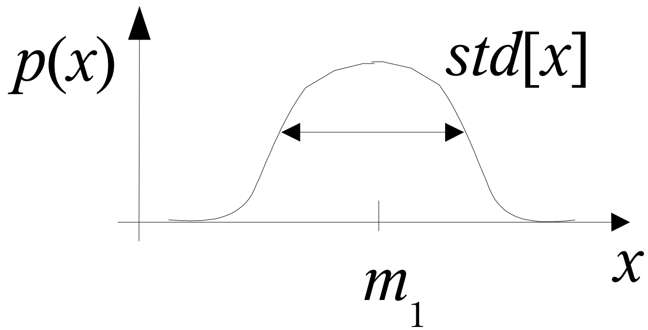
review

For non-zero mean more interesting is the **variance**, i.e. the power of the deviation from the mean.

$$var[X] = E[(X - m_1)^2] = E[X^2] - (E[X])^2$$

A metric for the spread around the mean is the **standard deviation**

$$std[X] = \sqrt{var[X]}$$



Random Variables - Poisson distribution

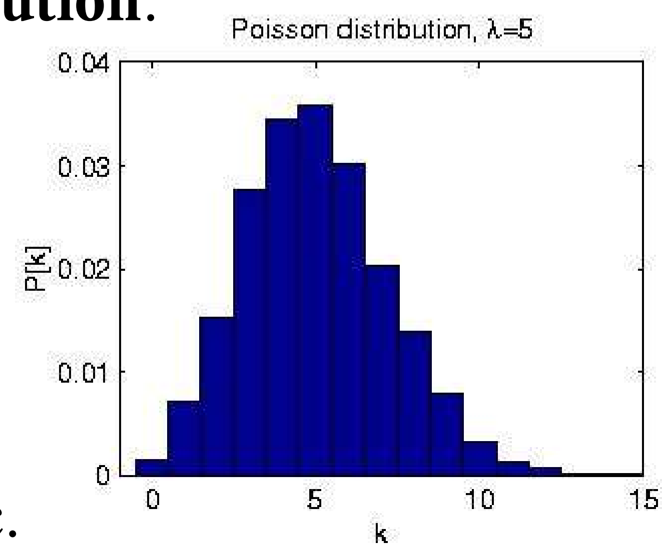
review

For a sequence of Bernoulli trials increment $N \rightarrow N+1$ with probability p starting at N=0. The probability of $N=k$ after n trials is given by *binomial distribution*.

$$P[N=k]=\binom{n}{k} p^k (1-p)^{n-k}$$

For large n and small p so that, $\lambda = np$, is moderate size this can be approximated by the **Poisson distribution**:

$$P[N=k]=\frac{\lambda^k e^{-\lambda}}{k!}$$



Typical examples are photon count in detector, spike counts, histogram values, etc.

Random Variables - Poisson distribution

review

The following moments are easy to compute using normalization

$$\sum_{k=0}^{\infty} P[k] = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1$$

Mean:
$$E[k] = \sum_{k=0}^{\infty} k\, P(k) = \lambda$$

Second moment:
$$E[k^2] = \sum_{k=0}^{\infty} k^2 P(k) = \lambda^2 + \lambda$$

Variance:
$$var[k] = E[k^2] - E[k]^2 = \lambda$$

The sum of two independent Poisson variables is also Poisson with

$$\lambda = \lambda_1 + \lambda_2$$

Radioactive decay – Poisson distribution

The number of radioactive decays, *k*, which occur within a time interval T ($\ll T_{1/2}$) is Poisson distributed

$$p(k|\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}$$

Here λ represents the mean number of decays during (t ... $t+T$).

$$\lambda=N(t)\ln 2\, T/T_{1/2}$$

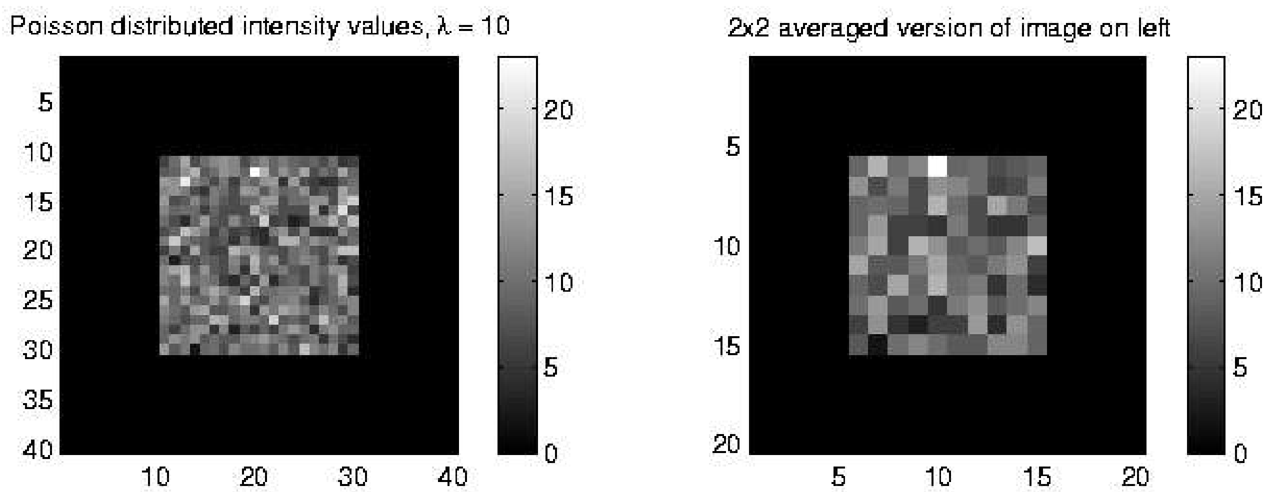
It can also be thought of as the **intensity** of the source.

Note that for the Poisson distribution the variance is also λ. The SNR of the source intensity is therefore:

$$SNR=\frac{E[k]}{std[k]}=\frac{\lambda}{\sqrt{\lambda}}=\sqrt{\lambda}$$

Radioactive decay – Poisson distribution

Assignment 7: Establish the relation between SNR and resolution for a Poisson distributed intensity. Do we gain in SNR as we reduce resolution with the same total number of radioactive decays?



Maximum Likelihood Estimation

Given measured data *y* and a PDF model $p(y|\theta)$ parameterized with model parameters θ the Maximum Likelihood estimation is

$$\hat{\theta} = \arg\max_{\theta} \; p(y|\theta)$$

The Maximum Likelihood estimate gives the

model that make the observation most likely.

We often find the ML estimate by solving for θ in:

$$\frac{\partial \ln p(y|\theta)}{\partial \theta} = 0$$

ML of Poisson observations

Assume you measure a Poisson process N times and get counts k_1, ..k_N. What is the parameter λ that explains those observations best. The ML estimate will be

$$\hat{\lambda} = \underset{\lambda}{arg\,max}\; p(k_{1,} \ldots, k_N | \lambda) = \underset{\lambda}{arg\,max}\; \log p(k_{1,} \ldots, k_N | \lambda)$$

The solution is given by the maximum of the Log-Likelihood:

$$\log L(\lambda) = \sum_{i=1}^{N} \log p(k_i | \lambda) = \sum_{i=1}^{N} \left(k_i \log \lambda - \lambda - \log k_i! \right)$$

Solving for $\partial \log L(\lambda) / \partial \lambda = 0$ gives, as one would expect:

$$\lambda = \frac{1}{N} \sum_{i=1} k_i$$

Maximum Likelihood Estimation

Each detector pair defines a line i. Denote the number of detected coincident events along each line as g_i, and the source intensity in pixel j as f_j. In this example the data is then $\mathbf{g} = [g_1, \ldots, g_N]$, and parameters are $\mathbf{f} = [f_1, \ldots, f_M]$. Since the detector readings are independent (given the data) the likelihood factors

$$p(\mathbf{g}|\mathbf{f}) = \prod_{i=1}^{N} p(g_i|\mathbf{f})$$

Each g is a count of independent radioactive decay events, and satisfies therefore a Poisson distribution,

$$p(g|\lambda) = \frac{\lambda^g}{g!} e^{-\lambda}$$

where $\lambda = E[g]$ is the mean or expected number of events.

Maximum Likelihood Estimation

Let p_{ij} be the probability that an event emitted at pixel j will be detected by detector pair i. The expected number of detected events can then be written as:

$$\lambda_i = E[g_i] = \sum_j p_{ij} f_j$$

This equation corresponds to the forward model. It is a matrix vector representation of

$$g(\phi, s) = \int_L dl\, f(x, y)$$

where (ϕ, s) is indexed by i and (x,y) by j. The advantage of this notation is that p_{ij} can be made to represent also

$$g(\phi, s) = \int_L dl\, f(x, y) \exp\left(-\int_L dl'\, \mu(x, y)\right)$$

where is known $\mu(x, y)$ from a separate CT image.

Maximum Likelihood Estimation - EM

Combining these definitions the total log-Likelihood is then

$$\log p(\boldsymbol{g}|\boldsymbol{f}) = \sum_{ij} g_i \log(p_{ij} f_j) - p_{ij} f_j$$

This Likelihood can be maximised with the following EM iteration (proof in Lange & Carson, 1984):

$$f_j^{(k+1)} = f_j^{(k)} \sum_i p_{ij} \frac{g_i}{\sum_n p_{in} f_n^{(k)}}$$

assuming that p_{ij} is properly normalised: $\sum_i p_{ij} = 1$

The iteration can be initialised with $\boldsymbol{f}^{(0)}=[1, ..., 1]$ and converges typically after 10-20 steps.

Maximum Likelihood Estimation - EM

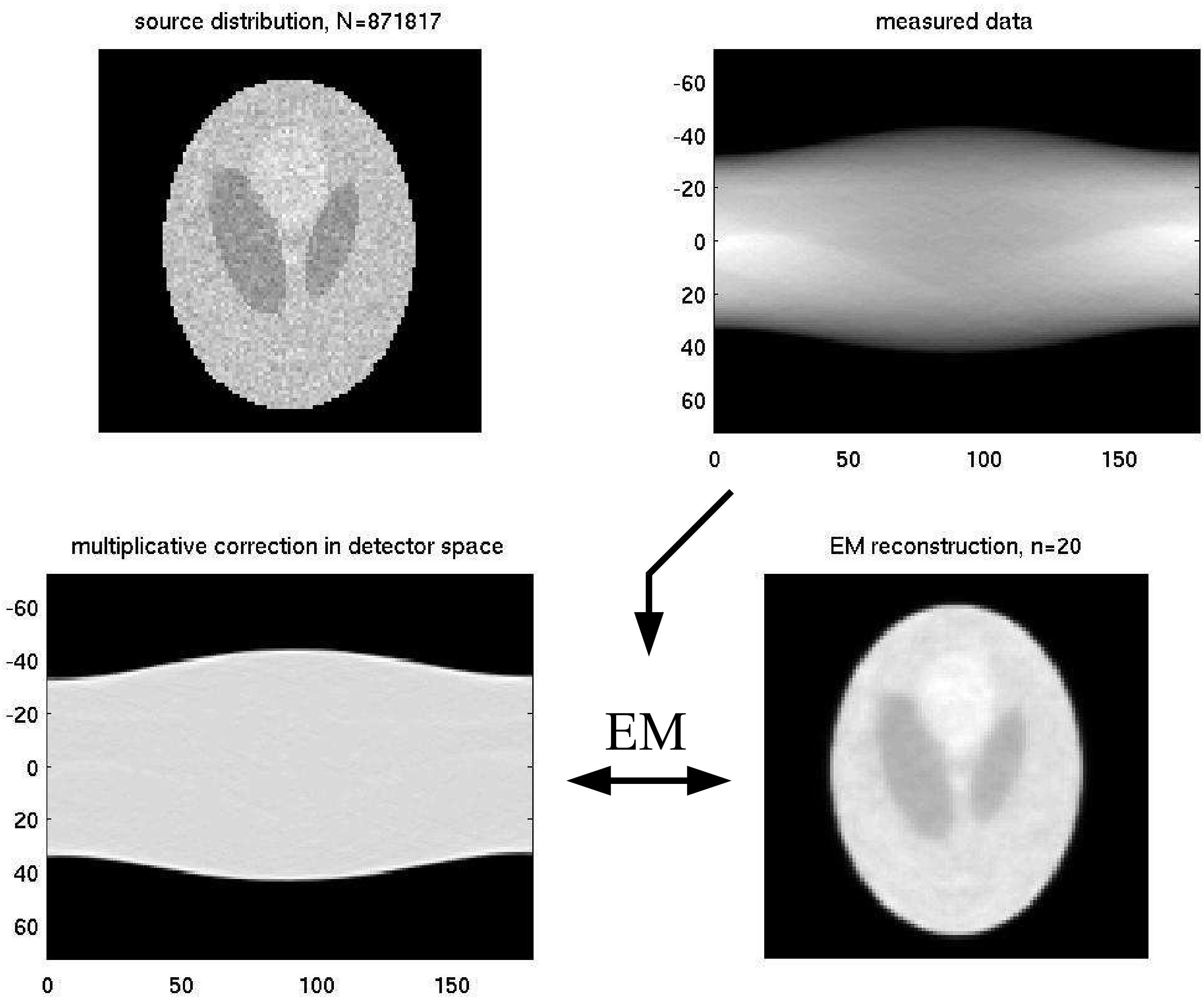
$$f_j^{(k+1)} = f_j^{(k)} \sum_i p_{ij} \frac{g_i}{\sum_n p_{in} f_n^{(k)}}$$

Note that $\sum_i p_{ij} ()_i$ corresponds to a back-projection operation

while $\sum_j p_{ij} ()_j$ is the forward projection (imaging).

Summary: The measured data is divided by the forward projection of the current estimate, i.e. the way the data would look if the current estimate was the real source. That ratio (which is constant =1 once the estimate is correct) is back-projected and multiplied with the current estimate. This gives an updated estimate. The process is repeated until the estimate no longer changes.

Maximum Likelihood Estimation - EM



Note that its constant within region of support

Maximum Likelihood Estimation - EM

Advantage over filtered back-projection:

- Can accommodate more general imaging scenarios
- Incorporates Poisson nature of data
  - The higher the intensity, the higher the noise.
  - Negative intensities not permitted.
- Easy to constrain region of support

Disadvantage:

- Slower by a factor of 10-20.