Stochastic Models of Objects and Images

Harrison H. Barrett

Center for Gamma-ray Imaging
University of Arizona

I. Basic concepts and notation

II. Some important probability laws

III. Random images

IV. Random objects

V. Random systems

VI. Challenges, unsolved problems, areas for future research

I. Basic concepts and notation

Mathematical descriptions of objects and images

- Objects are functions of continuous variables such as position or time:
  $f(\mathbf{r})$ or $f(\mathbf{r}, t)$, where $\mathbf{r} = (x, y, z)$

- Square-integrable functions can be regarded as vectors in a Hilbert space
  Notation: $f(\mathbf{r}) \Rightarrow \mathbf{f}$, vector in space $\mathbb{U}$.

- Hilbert spaces are linear vector spaces in which a scalar product is defined:
  Notation: $(\mathbf{f}_1, \mathbf{f}_2) = \mathbf{f}_1^\dagger \mathbf{f}_2 \equiv \int_{\mathbf{S}} d^3r \, f_1^*(\mathbf{r}) \, f_2(\mathbf{r})$

- Images *can* be functions of continuous variables ....
  ... but digital images are discrete sets of numbers:

$$\mathbf{g} = \{g_m, m = 1, ...M\}, \qquad M = \text{ total number of measurements}$$

- A digital image is a vector in a finite-dimensional Hilbert space $\mathbb{V}$

$$(\mathbf{g}_1, \mathbf{g}_2) = \mathbf{g}_1^\dagger \mathbf{g}_2 \equiv \sum_{m=1}^{M} g_{1m}^* \, g_{2m}$$

Deterministic (noise-free) descriptions of imaging systems

- A general imaging system maps object space to image space:

$$\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}\,, \qquad \boldsymbol{\mathcal{H}} : \mathbb{U} \to \mathbb{V}$$

- Many systems of interest are linear:

$$\boldsymbol{\mathcal{H}}\{\alpha\mathbf{f}_1 + \beta\mathbf{f}_2\} = \alpha\boldsymbol{\mathcal{H}}\{\mathbf{f}_1\} + \beta\boldsymbol{\mathcal{H}}\{\mathbf{f}_2\}\,, \qquad \alpha, \beta \text{ scalars}$$

- A general linear CD mapping of a 3D function to a discrete MD vector is defined by

$$g_m = [\boldsymbol{\mathcal{H}}\mathbf{f}]_m = \int_{\mathbf{S}} d^3r\, h_m(\mathbf{r})\, f(\mathbf{r})\,, \tag{7.225}$$

where $h_m(\mathbf{r})$ is the *point response function* (PRF) or *sensitivity function*.

Discrete representations of objects

Though real objects are functions in an infinite-dimensional Hilbert space, we often approximate them as finite sums of *expansion functions* such as pixels:

$$f_a(\mathbf{r}) = \sum_{n=1}^{N} \theta_n \, \phi_n(\mathbf{r}) \,, \tag{7.27}$$

where subscript a denote *approximate*

• The subscript is crucial; we must not confuse $f_a(\mathbf{r})$ and $f(\mathbf{r})$.

• The coefficients in the expansion, the set $\{\theta_n, \, n = 1, ..., N\}$, can be regarded as an N-dimensional vector denoted $\boldsymbol{\theta}$; if the expansion functions are known, this vector is an equivalent way of specifying the approximate representation $f_a(\mathbf{r})$.

Approximate matrix descriptions of linear imaging systems

Application of an imaging operator $\boldsymbol{\mathcal{H}}$ to the approximate object gives

$$\mathbf{g}_a = \boldsymbol{\mathcal{H}}\{f_a(\mathbf{r})\} = \sum_{n=1}^{N} \theta_n \boldsymbol{\mathcal{H}}\{\phi_n(\mathbf{r})\}\,. \tag{7.302}$$

For a linear CD mapping, the m^{th} component of $\mathbf{g}_a$ is given by

$$g_{am} = \sum_{n=1}^{N} \theta_n \int_{\mathbf{S}_f} d^q r\ h_m(\mathbf{r})\,\phi_n(\mathbf{r})\,, \qquad m = 1, ..., M\,. \tag{7.303}$$

Define an $M \times N$ matrix $\mathbf{H}$ with elements H_{mn} given by

$$H_{mn} = \int_{\mathbf{S}_f} d^q r\ h_m(\mathbf{r})\,\phi_n(\mathbf{r})\,, \tag{7.304}$$

so that

$$\mathbf{g}_a = \mathbf{H}\boldsymbol{\theta}\,. \tag{7.305}$$

Eq. (7.305) is a discrete-to-discrete (DD) model for an imaging system.

Definition of adjoint (backprojection)

Basic definition (Barrett and Myers, Sec. 1.3.7):

$$(\mathrm{g}_2, \boldsymbol{\mathcal{H}}\mathrm{f}_1)_{\mathbb{V}} = (\boldsymbol{\mathcal{H}}^\dagger \mathrm{g}_2, \mathrm{f}_1)_{\mathbb{U}} \tag{1.39}$$

Note that scalar product on the left is in image space, the one on the right is in object space

It is legal *by definition* to move an operator to the other side in a scalar product and replace it by its adjoint.

For a matrix operator:

$$\left[\mathbf{H}^\dagger \mathbf{g}\right]_n = \sum_{m=1}^{M} H^*_{mn} g_m \qquad \text{or} \qquad \left[\mathbf{H}^\dagger\right]_{nm} = [\mathbf{H}]^*_{mn}$$

For a CD operator,

$$\left[\mathbf{H}^\dagger \mathbf{g}\right](\mathbf{r}) = \sum_{m=1}^{M} h^*_m(\mathbf{r})\, g_m$$

The imaging equation

We often write real, noisy data as

$$\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f} + \mathbf{n}\,, \tag{15.10}$$

where $\boldsymbol{\mathcal{H}}$ is a linear CD operator $\mathbf{n}$ is an $M \times 1$ vector of "measurement noise" (not necessarily independent of $\mathbf{f}$). A simple mathematical tautology allows us to write

$$\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}_a + \boldsymbol{\mathcal{H}}\mathbf{f} - \boldsymbol{\mathcal{H}}\mathbf{f}_a + \mathbf{n} \equiv \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\epsilon}\,, \tag{15.11}$$

where the overall error $\boldsymbol{\epsilon}$ (modeling error plus noise) is given by

$$\boldsymbol{\epsilon} = \boldsymbol{\mathcal{H}}\mathbf{f} - \boldsymbol{\mathcal{H}}\mathbf{f}_a + \mathbf{n}\,, \tag{15.12}$$

The data vector $\mathbf{g}$ is random because:

- the measurement noise is random;
- the object is random;
- both $\boldsymbol{\mathcal{H}}$ and $\mathbf{H}$ can be random;
- the modeling error depends on the random object and system

Probability law for the data vector

An M-dimensional random vector g is a set of M scalar random variables, $\{g_m, m = 1, ..., M\}$. (Think of measurements from some detector array.)

If each g_m can take a continuous range of values, the vector g is described by a probability density function (PDF) $\mathrm{pr}(\mathrm{g}) = \mathrm{pr}(\{g_m\})$, and the mean of g is given by

$$\langle \mathrm{g} \rangle = \overline{\mathrm{g}} = \{\overline{g}_m\} = \int_{-\infty}^{\infty} dg_1 \int_{-\infty}^{\infty} dg_2 \cdots \int_{-\infty}^{\infty} dg_M \, \mathrm{g} \; \mathrm{pr}(\mathrm{g}) = \int_{\infty} d^M g \; \mathrm{g} \; \mathrm{pr}(\mathrm{g}) \, .$$

If each g_m can take on only integer values, the vector g is described by a probability (not PDF) $\mathrm{Pr}(\mathrm{g}) = \mathrm{Pr}(\{g_m\})$, and the mean of g is given by

$$\langle \mathrm{g} \rangle = \overline{\mathrm{g}} = \{\overline{g}_m\} = \sum_{g_1=0}^{\infty} \sum_{g_2=0}^{\infty} \cdots \sum_{g_M=0}^{\infty} \mathrm{g} \; \mathrm{Pr}(\mathrm{g}) \, .$$

(Note that the sum here is over each of the g_m, not over m.)

Covariance matrix

The *covariance matrix* is a generalization of the variance to random vectors.

For an MD random vector $\mathbf{g}$, the covariance matrix $\mathbf{K}$ is an $M \times M$ matrix with elements given by

$$K_{ij} = \left\langle (g_i - \overline{g}_i)(g_j - \overline{g}_j)^* \right\rangle \,, \tag{8.16}$$

where the asterisk indicates complex conjugate, allowing for the possibility that components of $\mathbf{g}$ might be complex. It follows from this definition that $\mathbf{K}$ is Hermitian, *i.e.*, $K_{ij} = K^*_{ji}$.

In outer-product form,

$$\mathbf{K} = \left\langle \, [\mathbf{g} - \overline{\mathbf{g}}] \, [\mathbf{g} - \overline{\mathbf{g}}]^\dagger \, \right\rangle \,.$$

Any random variable covaries with itself. The diagonal elements of the covariance matrix are the variances of the components:

$$K_{jj} = \mathrm{Var}\{g_j\} \,. \tag{8.18}$$

Characteristic functions

For any random variable or vector, the *characteristic function* is the expectation of the appropriate Fourier kernel:

$$\psi(\xi) \equiv \left\langle\, e^{-2\pi i \xi x} \,\right\rangle \,.$$

If x is scalar and real-valued, then

$$\psi(\xi) = \int_{-\infty}^{\infty} dx \;\mathrm{pr}(x)\, e^{-2\pi i \xi x} \,, \tag{C.53}$$

and the PDF and characteristic function form a Fourier transform pair:

$$\mathrm{pr}(x) = \int_{-\infty}^{\infty} d\xi \;\psi(\xi)\, e^{2\pi i \xi x} \,. \tag{C.54}$$

Caution: Do not confuse ξ with a spatial frequency.

Moments of the random variable x can be derived through differentiation of $\psi(\xi)$:

$$\left\langle\, x^k \,\right\rangle = (-2\pi i)^{-k} \left.\frac{\partial^k}{\partial \xi^k} \psi(\xi) \right|_{\xi=0} \,. \tag{C.55}$$

Characteristic function of a real random vector

For a real $M \times 1$ random vector g (column vector), the characteristic function is defined as

$$\psi_{\mathrm{g}}(\boldsymbol{\xi}) = \left\langle \exp(-2\pi i \boldsymbol{\xi}^t \mathrm{g}) \right\rangle \,, \tag{8.26}$$

where $\boldsymbol{\xi}^t$ is a real $1 \times M$ vector.

For the case of a continuous-valued random vector, $\psi_{\mathrm{g}}(\boldsymbol{\xi})$ can be written as

$$\psi_{\mathrm{g}}(\boldsymbol{\xi}) = \int_{\infty} d^M g \; \mathrm{pr}(\mathrm{g}) \exp(-2\pi i \boldsymbol{\xi}^t \mathrm{g}) \,. \tag{8.27}$$

This integral is the MD Fourier transform of the PDF, so

$$\mathrm{pr}(\mathrm{g}) = \int_{\infty} d^M \xi \; \psi_{\mathrm{g}}(\boldsymbol{\xi}) \exp(2\pi i \boldsymbol{\xi}^t \mathrm{g}) \,. \tag{8.28}$$

Linear transformations of random vectors

If the random vector $\mathbf{g}$ is generated as the output of a linear filter acting on the random vector $\mathbf{f}$, we can characterize the linear transformation by an $M \times N$ matrix $\mathbf{H}$. Then we can write the $M \times 1$ output vector $\mathbf{g}$ in terms of the $N \times 1$ input vector $\mathbf{f}$ as

$$\mathbf{g} = \mathbf{H}\mathbf{f}\,. \tag{8.40}$$

From the linearity of the expectation operator, we have immediately for the mean of $\mathbf{g}$,

$$\overline{g} = \langle \mathbf{g} \rangle = \langle \mathbf{H}\mathbf{f} \rangle = \mathbf{H}\langle \mathbf{f} \rangle = \mathbf{H}\overline{\mathbf{f}}\,. \tag{8.49}$$

The covariance matrix of $\mathbf{g}$ is found as

$$\begin{aligned} \mathbf{K}_\mathbf{g} &= \left\langle \Delta\mathbf{g}\Delta\mathbf{g}^\dagger \right\rangle = \left\langle (\mathbf{H}\mathbf{f} - \mathbf{H}\overline{\mathbf{f}})(\mathbf{H}\mathbf{f} - \mathbf{H}\overline{\mathbf{f}})^\dagger \right\rangle \\ &= \mathbf{H}\left\langle \Delta\mathbf{f}\,\Delta\mathbf{f}^\dagger \right\rangle \mathbf{H}^\dagger = \mathbf{H}\mathbf{K}_\mathbf{f}\mathbf{H}^\dagger\,, \end{aligned} \tag{8.50}$$

where $\Delta\mathbf{f} \equiv \mathbf{f} - \overline{\mathbf{f}}$.

Linear transformation of the characteristic function

Transformation of the PDF is tricky, but the characteristic function is easy:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \left\langle \exp(-2\pi i \boldsymbol{\xi}^t \mathbf{H}\mathbf{f}) \right\rangle = \left\langle \exp\left[-2\pi i (\mathbf{H}^t\boldsymbol{\xi})^t \mathbf{f}\right] \right\rangle , \tag{8.42}$$

where the last step has used the definition of the adjoint, (1.39). Thus

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \psi_{\mathbf{f}}(\mathbf{H}^t\boldsymbol{\xi}) , \tag{8.43}$$

so knowledge of $\psi_{\mathbf{f}}$ and $\mathbf{H}$ immediately gives $\psi_{\mathbf{g}}$.

The PDF on $\mathbf{g}$ can in principle be found by taking an inverse MD Fourier transform of (8.43). Formally, we can write

$$\mathsf{pr}(\mathbf{g}) = \int_\infty d^M\xi \; \psi_{\mathbf{f}}(\mathbf{H}^t\boldsymbol{\xi}) \exp(2\pi i \boldsymbol{\xi}^t \mathbf{g}) , \tag{8.44}$$

but in practice the integral might not be easy. The problem is that we are integrating a function of an ND vector over an MD space.

Intro to random processes

Simple definition: A random process is a random variable that varies with space and/or time. More formally, a spatial random process is a function of two variables, $\mathbf{r}$ and ζ. Depending on the context, $f(\mathbf{r}, \zeta)$ can refer to:

1. The family of spatial functions, referred to as the ensemble; in this case, $\mathbf{r}$ and ζ are variables (e.g., all possible images)

2. A single realization or sample of the spatial functions; in this case, $\mathbf{r}$ is variable and ζ is fixed (one particular image)

3. The random variable at a single point; in this case, $\mathbf{r}$ is fixed and ζ is variable (value at one point in all possible images)

4. A single number; in this case, $\mathbf{r}$ is fixed and ζ is fixed (value at one point in one image)

Notational quirks: Usually drop the ζ designator and hope the differences above will be clear by context

Averages

Consider a scalar-valued continuous random process. For fixed $\mathbf{r}$, $f(\mathbf{r})$ is simply a random variable (interpretation 3), and its expectation is defined just as for any other random variable;

$$\mathrm{E}\{f(\mathbf{r})\} = \langle f(\mathbf{r}) \rangle = \bar{f}(\mathbf{r}) = \int_{-\infty}^{\infty} df(\mathbf{r}) \; f(\mathbf{r}) \, \mathrm{pr}[f(\mathbf{r})] \, . \tag{8.71}$$

Computation of this expectation requires only the univariate PDF $\mathrm{pr}[f(\mathbf{r})]$.

N.B. The integral is over $f(\mathbf{r})$, not $\mathbf{r}$; result can still be a function of $\mathbf{r}$.

Multiple-point expectations

$$\langle\, f(\mathbf{r}_1) f(\mathbf{r}_2)\,\rangle = \int_{-\infty}^{\infty} df(\mathbf{r}_1) \int_{-\infty}^{\infty} df(\mathbf{r}_2)\;\; f(\mathbf{r}_1)\, f(\mathbf{r}_2)\, \mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]\,. \tag{8.74}$$

Here, $f(\mathbf{r}_1)$ and $f(\mathbf{r}_2)$ must be regarded as two *distinct* random variables and $\mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]$ is their joint density. Only in very special circumstances will it be possible to write $\mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]$ as $\mathrm{pr}[f(\mathbf{r}_1)]\,\mathrm{pr}[f(\mathbf{r}_2)]$.

A general two-point moment is defined by

$$\langle\, [f(\mathbf{r}_1)]^m\, [f(\mathbf{r}_2)]^n\,\rangle =$$

$$\int_{-\infty}^{\infty} df(\mathbf{r}_1) \int_{-\infty}^{\infty} df(\mathbf{r}_2)\; [f(\mathbf{r}_1)]^m\, [f(\mathbf{r}_2)]^n\, \mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2)]\,. \tag{8.75}$$

Any moment involving the K points $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_K$ can be computed if $\mathrm{pr}[f(\mathbf{r}_1), f(\mathbf{r}_2), ..., f(\mathbf{r}_K)]$ is known. If this K-fold joint density is known for all values of each of the $\mathbf{r}_k$, the process is said to be *fully characterized* to order K.

Correlation analysis

The autocorrelation function $R(\mathbf{r}_1, \mathbf{r}_2)$ of a spatial random process $f(\mathbf{r})$ is defined by

$$R(\mathbf{r}_1, \mathbf{r}_2) = \langle f(\mathbf{r}_1)\, f^*(\mathbf{r}_2) \rangle \ . \tag{8.97}$$

The autocovariance function $K(\mathbf{r}_1, \mathbf{r}_2)$ is the continuous analog of the covariance matrix:

$$\begin{aligned} K(\mathbf{r}_1, \mathbf{r}_2) &= \big\langle [f(\mathbf{r}_1) - \langle f(\mathbf{r}_1) \rangle] \left[f^*(\mathbf{r}_2) - \langle f^*(\mathbf{r}_2) \rangle \right] \big\rangle \\ &= R(\mathbf{r}_1, \mathbf{r}_2) - \bar{f}(\mathbf{r}_1)\, \bar{f}^*(\mathbf{r}_2) \, . \end{aligned} \tag{8.98}$$

The autocovariance function reduces to the variance when $\mathbf{r}_2 = \mathbf{r}_1 = \mathbf{r}$, *i.e.*,

$$K(\mathbf{r}, \mathbf{r}) = R(\mathbf{r}, \mathbf{r}) - |\bar{f}(\mathbf{r})|^2 = \mathrm{Var}\{ f(\mathbf{r}) \} \, . \tag{8.99}$$

Temporal stationarity

Stationarity in the strict sense requires that all K-point PDFs be independent of the time origin.

Stationarity in the wide (or loose) sense requires only that the mean and autocorrelation have no preferred origin. For a temporal random process $f(t)$, this means:

$$\langle f(t) \rangle = \text{constant} ;$$

$$R(t_1, t_2) = R(t_1 - t_2) .$$

Electrical noise and random optical fields are temporally stationary if the parameters describing the noise don't change in time. For example, electrical noise from a resistor is temporally stationary if temperature $=$ const.

A stationary random process is said to be *ergodic in the mean* if a time average gives the (constant) ensemble average. Stationary Gaussian random processes are ergodic.

Spatial stationarity?

Strict-sense and wide-sense stationarity can be defined for spatial random processes just as for temporal ones, but the conditions are much harder to satisfy.

- Snow on a TV screen is stationary if you ignore the raster
- Poisson noise in an image is stationary for a uniform grey scene

My advice: Never assume spatial stationarity in imaging problems

Possible exception: Well, you might get away with assuming that atmospheric turbulence is stationary.

Spectral analysis

For a stationary temporal random process, we can follow Khinchin and define the *power spectrum* or *power spectral density* as

$$S(\nu) = \mathcal{F}\{R(\Delta t)\} = \int_{-\infty}^{\infty} d\Delta t \ \langle f(t+\Delta t)\, f^*(t)\rangle \exp(-2\pi i \nu \Delta t)\,. \tag{8.133}$$

This equation is often called the Wiener-Khinchin theorem, but I like to think of it as a definition. (Besides, Wiener didn't consider ensemble averages at all.)

What to do if the process is not stationary? Use the same equation anyway! (You just can't assume that the lefthand side is a function only of ν.)

And if you do that, you can use spatial as well as temporal random processes.

Stochastic Wigner distribution function

Rewrite the spatial autocorrelation function in symmetrized form:

$$R(\mathbf{r}_1, \mathbf{r}_2) \rightarrow R(\mathbf{r}_0, \Delta\mathbf{r})$$

where

$$\mathbf{r}_0 = \tfrac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)\,, \qquad \Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2\,. \tag{8.120}$$

There is no loss of generality if we don't assume the function factors; $R(\mathbf{r}_0, \Delta\mathbf{r})$ is still a general function of four variables if $\mathbf{r}$ is 2D.

Now take a FT with respect to $\Delta\mathbf{r}$ only, and define the *stochastic Wigner distribution function* as

$$W_f(\mathbf{r}_0, \boldsymbol{\rho}) = \int_\infty d^q\Delta r \;\left\langle f(\mathbf{r}_0 + \tfrac{1}{2}\Delta\mathbf{r})\, f^*(\mathbf{r}_0 - \tfrac{1}{2}\Delta\mathbf{r})\right\rangle \exp(-2\pi i\boldsymbol{\rho}\cdot\Delta\mathbf{r})\,. \tag{8.140}$$

If (somehow), $f(\mathbf{r})$ really was stationary, $W_f(\mathbf{r}_0, \boldsymbol{\rho})$ would be independent of $\mathbf{r}_0$ and equal to the spatial power spectrum.

PDFs and characteristic functionals

A spatial random process $f(\mathbf{r})$ is an infinite-dimensional vector in a Hilbert space – *if* every sample function is square-integrable.

Thus, at best, an infinite-dimensional PDF would be needed for a complete characterization: very tricky mathematically.

BUT, an infinite-dimensional characteristic *functional* can always be defined – and often calculated explicitly!

ALL statistical properties of a random process are contained in its characteristic functional.

Characteristic functionals – definition

Recall the definition of the characteristic *function* for a MD real random vector:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \left\langle \exp(-2\pi i \boldsymbol{\xi}^t \mathbf{g}) \right\rangle \,, \tag{8.26}$$

Here, $\boldsymbol{\xi}$ is a real $M \times 1$ vector, and $\boldsymbol{\xi}^t \mathbf{g}$ denotes a scalar product.

In the case of a random process $f(\mathbf{r})$, each sample function corresponds to a vector $\mathbf{f}$ in an infinite-dimensional Hilbert space, so the frequency vector $\boldsymbol{\xi}$ in (8.26) must be replaced by an infinite-dimensional vector $\mathbf{s}$ in the same Hilbert space as $\mathbf{f}$. That means that $\mathbf{s}$ describes a function $s(\mathbf{r})$, so the characteristic function becomes a characteristic *functional* $\Psi_{\mathbf{f}}\{s(\mathbf{r})\}$ or $\Psi_{\mathbf{f}}(\mathbf{s})$ for short. It is defined by

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \langle \exp[-2\pi i (\mathbf{s}, \mathbf{f})] \rangle \,, \tag{8.94}$$

where $(\mathbf{s}, \mathbf{f})$ is the usual $\mathbb{L}_2$ scalar product.

Linear transformations of random processes

Consider the familiar form of a general linear mapping, $\mathrm{g} = \boldsymbol{\mathcal{H}}\mathrm{f}$. If $\boldsymbol{\mathcal{H}}$ is a CD mapping, then

$$g_m = \int d^q r \, h_m(\mathbf{r}) \, f(\mathbf{r}) \, .$$

If f denotes a random process, this mapping defines a random vector. Specifically, if one sample function of $f(\mathbf{r})$ is denoted $f(\mathbf{r}, \zeta)$, then $\mathrm{g} = \boldsymbol{\mathcal{H}}\mathrm{f}$ is to be interpreted as

$$g_m(\zeta) = \int d^q r \, h_m(\mathbf{r}) \, f(\mathbf{r}, \zeta) \, .$$

A similar interpretation applies to integral transforms. For example, the Fourier transform of a random process is a new random process obtained by taking the Fourier transform of each sample function.

Linear transformation of the characteristic functional

Back to the definition, with a different notation for the scalar product:

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \left\langle \exp[-2\pi i\, \mathbf{s}^\dagger \mathbf{f}] \right\rangle \,, \tag{8.94}$$

For now, assume $\mathbf{s}$ and $\mathbf{f}$ are real.

Consider the general linear mapping $\mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}$. The characteristic *function* of the random vector $\mathbf{g}$ is given by

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) \equiv \left\langle \exp[-2\pi i\, \boldsymbol{\xi}^\dagger\, \mathbf{g}\,] \right\rangle = \left\langle \exp[-2\pi i\, \boldsymbol{\xi}^\dagger (\boldsymbol{\mathcal{H}}\mathbf{f})] \right\rangle = \left\langle \exp[-2\pi i\, (\boldsymbol{\mathcal{H}}^\dagger \boldsymbol{\xi})^\dagger \mathbf{f}] \right\rangle \,,$$

where the last step follows from the definition of the adjoint. Thus

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}(\boldsymbol{\mathcal{H}}^\dagger \boldsymbol{\xi}\,) \,. \tag{8.96}$$

Very powerful result: If we know the characteristic *functional* for $\mathbf{f}$, we immediately have the characteristic *function* for $\boldsymbol{\mathcal{H}}\mathbf{f}$, and therefore we know *all statistical properties* of the random vector (e.g., image) $\mathbf{g}$.

II. Some important probability laws

Poisson and Gaussian statistics

- The Poisson law
  - Applies to integer-valued discrete random variables
  - Often arises from independence assumptions
- The normal or Gaussian law
  - Applies to continuous random variables
  - Often arises as result of central-limit theorem

Both are crucial in imaging

Both are readily generalized to random vectors or random processes.

The univariate Poisson law

If one counts identical, indistinguishable events (e.g. photons, it is found that the probability of observing n events is given by

$$\mathrm{Pr}(n) = \frac{e^{-\lambda}\,\lambda^n}{n!}\,, \qquad n = 0, 1, 2, \ldots\,. \tag{C.165}$$

Note that this probability law is specified entirely by the single number λ, often called *the parameter* of the Poisson distribution.

The parameter λ is also the mean value of n:

$$\langle n \rangle = \sum_{n=0}^{\infty} n\,\mathrm{Pr}(n) = \lambda\,. \tag{C.167}$$

Moreover, λ is also the variance.

The multivariate Poisson law

Ref.: B&M Chap. 11

If an array of J detectors counts independent photons, the multivariate probability law of the outputs must have the form,

$$\Pr(\{g_j\}) = \prod_{j=1}^{J} \exp(-\overline{g}_j) \, \frac{(\overline{g}_j)^{g_j}}{g_j!} \, . \tag{11.40}$$

Because of this product form, counts in different elements are statistically independent and hence uncorrelated.

Also, the variance of the counts in one element is equal to its mean, so we can write the covariance matrix elements as

$$K_{jk} = \left\langle [g_j - \overline{g}_j] \, [g_k - \overline{g}_k] \right\rangle = \overline{g}_j \, \delta_{jk} \, . \tag{11.41}$$

Poisson random processes

Temporal and spatial Poisson random processes are defined by

$$z(t) = \sum_{n=1}^{N} \delta(t - t_n)\,, \qquad (0 < t \leq T) \tag{11.64}$$

$$g(\mathbf{r}) = \sum_{n=1}^{N} \delta(\mathbf{r} - \mathbf{r}_n), \qquad (\mathbf{r} \in A) \tag{11.72}$$

Conditions for these sums to be Poisson random processes are discussed in Sec. 11.3. Roughly, they require that the events just be independent. In that case the number of events N in time interval T or area A is a Poisson random variable.

Properties of spatial Poisson random processes

$$\mathrm{Pr}(1 \text{ event in } \Delta A) = b(\mathbf{r})\,\Delta A\,. \tag{11.73}$$

$$\mathrm{pr}(\{\mathbf{r}_n\}|N) = \prod_{n=1}^{N} \mathrm{pr}(\mathbf{r}_n)\,; \tag{11.75}$$

$$\mathrm{pr}(\mathbf{r}_n) = \frac{b(\mathbf{r}_n)}{\int_A d^2r\ b(\mathbf{r})}\,; \tag{11.76}$$

$$\mathrm{Pr}(N) = \frac{\overline{N}^N}{N!}\,\mathrm{exp}(-\overline{N})\,; \tag{11.77}$$

$$\overline{N} = \int_A d^2r\ b(\mathbf{r})\,. \tag{11.78}$$

$$\mathrm{E}\{g(\mathbf{r})\} = \overline{N}\,\mathrm{pr}(\mathbf{r}) = b(\mathbf{r})\,. \tag{11.84}$$

Properties of spatial Poisson random processes – cont.

All statistical properties of $g(\mathbf{r})$ are determined by the *photon fluence* $b(\mathbf{r})$, provided that fluence is not random.

In particular, the characteristic functional is given by

$$\Psi_{\mathbf{g}}(\mathbf{s}) = \exp\left\{-\overline{N} + \int_A d^2 r_n \; b(\mathbf{r}_n)\, e^{-2\pi i \, s(\mathbf{r_n})}\right\} , \qquad (11.150)$$

For a stationary Poisson random process where $b(\mathbf{r})$ is constant over an area A, the fluence is $\overline{N}/A$, and we have

$$\Psi_{\mathbf{g}}(\mathbf{s}) = \exp\left\{-\overline{N}\left[1 + \frac{1}{A}\int_A d^2 r_n \; e^{-2\pi i \, s(\mathbf{r_n})}\right]\right\} . \qquad (11.151)$$

Univariate and multivariate Gaussian PDFs

The standard form of the Gaussian PDF for a scalar random variable is

$$\mathrm{pr}(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{(x-\overline{x})^2}{2\sigma^2}\right] . \qquad \text{(C.108)}$$

This is a two-parameter PDF, with $\overline{x}$ being the mean and σ^2 being the variance, as the notation implies.

The corresponding multivariate form for an $M \times 1$ random vector is

$$\mathrm{pr}(\mathbf{g}) = \left[(2\pi)^M \det(\mathbf{K})\right]^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{g}-\overline{\mathbf{g}})^t\, \mathbf{K}^{-1}(\mathbf{g}-\overline{\mathbf{g}})\right] , \qquad \text{(8.185)}$$

where $\overline{\mathbf{g}}$ is the mean vector and $\mathbf{K}$ is the covariance matrix of $\mathbf{g}$.

The multivariate characteristic function is the MD Fourier transform of this PDF:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \exp(-2\pi i \boldsymbol{\xi}^t \overline{\mathbf{g}}) \exp\left(-2\pi^2 \boldsymbol{\xi}^t \mathbf{K} \boldsymbol{\xi}\right) . \qquad \text{(8.196)}$$

N. B.: No inverse or determinant required!

Central limit theorem

Univariate form:

The PDF of a sum of N independent scalar random variables of (almost) arbitrary PDF tends to a univariate normal as $N \to \infty$. (In practice $N \simeq 5$ is often "good enough".)

Since the variables are independent, the mean of the sum is the sum of the means, and the variance of the sum is the sum of the variances, so the PDF on the sum is fully determined in the limit.

Multivariate form:

The PDF of a sum of N independent $M \times 1$ random vectors of (almost) arbitrary PDF tends to a multivariate normal as $N \to \infty$.

Gaussian random processes

A spatial random process $f(\mathbf{r})$ is said to be Gaussian if all multipoint densities and linear transformations of it are Gaussian.

An unconventional but useful definition is that $f(\mathbf{r})$ is Gaussian if its characteristic functional has the form:

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \exp(-2\pi i \mathbf{s}^\dagger \bar{\mathbf{f}}) \exp(-2\pi^2 \mathbf{s}^\dagger \mathcal{K}_{\mathbf{f}} \mathbf{s}) , \qquad (8.216)$$

where $\mathcal{K}_{\mathbf{f}}$ is the autocovariance operator, *i.e.*, the integral operator with kernel $K_{\mathbf{f}}(\mathbf{r}, \mathbf{r}')$.

All statistical properties of $f(\mathbf{r})$ are determined by its mean and autocovariance function. Again, no inverse or determinant is required.

III. Random images

Likelihoods: conditional image statistics

- Descriptive likelihoods

$$\text{pr}(\mathbf{g}|\mathbf{f}), \qquad \text{pr}(\hat{\mathbf{f}}|\mathbf{f}),$$
$$(\hat{\mathbf{f}} = \text{reconstructed image})$$

- Likelihoods for classification problems

$$\text{pr}(\mathbf{g}|H_j), \qquad \text{pr}(\hat{\mathbf{f}}|H_j),$$
$$(H_j = \text{ hypothesis that object is in class } C_j)$$

- Likelihoods for estimation problems

$$\text{pr}(\mathbf{g}|\boldsymbol{\theta}), \qquad \text{pr}(\hat{\mathbf{f}}|\boldsymbol{\theta})$$
$$(\boldsymbol{\theta} = \text{ parameter to be estimated})$$

- Likelihoods for joint detection and estimation

$$\text{pr}(\mathbf{g}|H_j, \boldsymbol{\theta}), \qquad \text{pr}(\hat{\mathbf{f}}|H_j, \boldsymbol{\theta})$$

Descriptive likelihoods

• Goal: Specify the conditional image statistics *for a specific object and a specific imaging system*

• Only remaining randomness is measurement noise, possibly propagated through a processing or reconstruction algorithm

• Measurement noise on output of integrating detectors is usually Gaussian

• Measurement noise on output of photon-counting detectors is Poisson for a specified object and system

• Noise in a processed image is more complicated

Conditional probability law for raw data in SPECT

Let $M \times 1$ vector $\mathbf{g}$ describe the raw projection data in SPECT
($M = \#$ projection angles $\times$ $\#$ of detector elements per projection)

$$\overline{g}_m = [\boldsymbol{\mathcal{H}}\mathbf{f}]_m = \int_{\mathbf{S}} d^3r \, h_m(\mathbf{r}) \, f(\mathbf{r})$$

Rigorously,

$$\Pr(\mathbf{g}|\mathbf{f}) = \prod_{m=1}^{M} \frac{(\overline{g}_m)^{g_m}}{g_m!} \exp(-\overline{g}_m)$$

This Poisson law is exact even when we consider:

- Detector quantum efficiency
- Scatter in object or detector
- Background radiation
- Position-estimation algorithm, energy window

(Follows from binomial selection theorem if each event is assigned to at most one bin in the data set.)

Conditional probability law for raw data in PET

Multivariate Poisson still works for events assigned to lines of response by any algorithm, but ....

... correction for accidental coincidences can cause non-Poisson behavior

Gaussian approximation to the Poisson

If the noise is Poisson but $\overline{g}_m >> 1$ for all m, a useful approximation is

$$\text{pr}(\mathbf{g}|\mathbf{f}) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi}\,\overline{g}_m} \exp\left[-\frac{(g_m - \overline{g}_m)^2}{2\overline{g}_m}\right]$$

Note that the measurements are independent but not identically distributed.

Effect of linear image reconstruction

Consider a linear reconstruction algorithm described by

$$\hat{\mathbf{f}} = \mathcal{O}\mathbf{g}$$

Transformed mean and covariance are:

$$\left\langle \hat{\mathbf{f}} \right\rangle_{\mathbf{g}|\mathbf{f}} = \mathcal{O}\mathcal{H}\mathbf{f}\,, \qquad \mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}} = \mathcal{O}\mathbf{K}_{\mathbf{g}|\mathbf{f}}\mathcal{O}^{\dagger}$$

If $\mathrm{pr}(\mathbf{g}|\mathbf{f})$ is Gaussian, then $\hat{\mathbf{f}}$ is also Gaussian (any linear transformation of a Gaussian is a Gaussian), and the PDF $\mathrm{pr}(\mathbf{g}|\mathbf{f})$ is fully determined by this mean and covariance.

Effect of linear image reconstruction – cont.

From last slide:

$$\left\langle \hat{\mathbf{f}} \right\rangle_{\mathbf{g}|\mathbf{f}} = \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{H}}\mathbf{f}\,, \qquad \mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}} = \boldsymbol{\mathcal{O}}\mathbf{K}_{\mathbf{g}|\mathbf{f}}\boldsymbol{\mathcal{O}}^{\dagger}$$

If the data are Poisson, then as we saw earlier,

$$\left[\mathbf{K}_{\mathbf{g}|\mathbf{f}}\right]_{mm'} = \overline{g}_m\,\delta_{mm'} = [\boldsymbol{\mathcal{H}}\mathbf{f}]_m\;\delta_{mm'}$$

Thus, for a discrete reconstruction,

$$\left[\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}\right]_{nn'} = \sum_{m=1}^{M} O_{nm}\,O_{n'm}\,[\boldsymbol{\mathcal{H}}\mathbf{f}]_m$$

Easy to compute, linear transformation of the object.

Since many independent data values contribute to each voxel, it is often an excellent approximation to assume that the linear reconstruction is Gaussian with this covariance matrix.

Effect of nonlinear image reconstruction

For nonlinear reconstruction algorithms, it is harder to to compute the descriptive likelihoods, but algorithms exist for:

• Computing pr $\left(\hat{\mathbf{f}}^{(k)}|\mathbf{f}\right)$ at each iteration k, of iterative algorithms such as MLEM and OSEM (Barrett, Wilson and Tsui, 1994)

• Computing pr $\left(\hat{\mathbf{f}}^{(\infty)}|\mathbf{f}\right)$ for MAP-like algorithms run to convergence (Fessler, 1996)

Often a good approximation to pr$(\hat{\mathbf{f}}\,|\mathbf{f}\,)$ for nonlinear algorithms with positivity is a multivariate log-normal.

For more discussion and references, see Chap. 15.

Likelihoods for classification problems

To perform a classification based on raw data $\mathbf{g}$, we need $\mathrm{pr}(\mathbf{g}|H_j)$, where $H_j =$ hypothesis that object is in class C_j

To perform a classification based on a reconstruction, we need $\mathrm{pr}(\hat{\mathbf{f}}\,|H_j)$.

Formally we can write

$$\mathrm{pr}(\mathbf{g}|H_j) = \int d\mathbf{f}\ \mathrm{pr}(\mathbf{g}|\mathbf{f})\,\mathrm{pr}(\mathbf{f}\,|H_j)\,,$$

where $\mathrm{pr}(\mathbf{g}|\mathbf{f})$ is the descriptive likelihood and $\mathrm{pr}(\mathbf{f}\,|H_j)$ is a huge-dimensional PDF for objects in class C_j (e.g., for a set of parameters that describe the objects).

Likelihoods for classification problems – cont.

From last slide,

$$\mathrm{pr}(\mathbf{g}|H_j) = \int d\mathbf{f}\ \mathrm{pr}(\mathbf{g}|\mathbf{f})\,\mathrm{pr}(\mathbf{f}|H_j)\,.$$

An equivalent notation is

$$\mathrm{pr}(\mathbf{g}|H_j) = \langle\, \mathrm{pr}(\mathbf{g}|\mathbf{f})\,\rangle_{\mathbf{f}|H_j}\ .$$

If we have real or simulated sample objects from class C_j, we can approximate this density by a sample average. (This point to be discussed further by Matt Kupinski.)

Doubly stochastic averaging for classification problems

Often we can perform averages with respect to pr($\mathbf{g}|H_j$) without knowing explicit functional forms.

Let $\mathbf{T}(\mathbf{g})$ be some arbitrary function of the raw data. Its average is

$$\langle\, \mathbf{T}(\mathbf{g}) \,\rangle_{\mathbf{g}|H_j} = \langle\langle\, \mathbf{T}(\mathbf{g}) \,\rangle_{\mathbf{g}|\mathbf{f}} \,\rangle_{\mathbf{f}|H_j} \equiv \overline{\overline{\mathbf{T}}}_j$$

Double overbar denotes average with respect to both measurement noise and object randomness

Can also define single-bar average over just measurement noise:

$$\overline{\mathbf{T}}(\mathbf{f}) \equiv \langle\, \mathbf{T}(\mathbf{g}) \,\rangle_{\mathbf{g}|\mathbf{f}}$$

Doubly stochastic covariance matrix

Covariance matrix on $\mathbf{g}$ with both measurement noise and object variability:

$$\mathbf{K_g} = \langle\langle \left[\mathbf{g} - \overline{\overline{\mathbf{g}}}\right] \left[\mathbf{g} - \overline{\overline{\mathbf{g}}}\right]^t \rangle_{\mathbf{g}|\mathbf{f}}\rangle_{\mathbf{f}} \,. \tag{8.344}$$

Adding and subtracting $\overline{\mathbf{g}}(\mathbf{f})$ in each factor gives

$$\mathbf{K_g} = \langle\langle \left[\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f}) + \overline{\mathbf{g}}(\mathbf{f}) - \overline{\overline{\mathbf{g}}}\right] \left[\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f}) + \overline{\mathbf{g}}(\mathbf{f}) - \overline{\overline{\mathbf{g}}}\right]^t \rangle_{\mathbf{g}|\mathbf{f}}\rangle_{\mathbf{f}} \,. \tag{8.345}$$

Since $\langle[\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f})]\rangle_{\mathbf{g}|\mathbf{f}} = 0$, we see that

$$\mathbf{K_g} = \langle\langle[\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f})][\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f})]^t\rangle_{\mathbf{g}|\mathbf{f}}\rangle_{\mathbf{f}} + \langle \left[\overline{\mathbf{g}}(\mathbf{f}) - \overline{\overline{\mathbf{g}}}\right] \left[\overline{\mathbf{g}}(\mathbf{f}) - \overline{\overline{\mathbf{g}}}\right]^t \rangle_{\mathbf{f}} \,. \tag{8.346}$$

Doubly stochastic covariance matrix – cont.

From last slide:

$$\mathbf{K_g} = \langle \langle [\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f})][\mathbf{g} - \overline{\mathbf{g}}(\mathbf{f})]^t \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}} + \langle \left[\overline{\mathbf{g}}(\mathbf{f}) - \overline{\overline{\mathbf{g}}}\right] \left[\overline{\mathbf{g}}(\mathbf{f}) - \overline{\overline{\mathbf{g}}}\right]^t \rangle_{\mathbf{f}} . \quad (8.346)$$

• First term in this expression is just the noise covariance matrix $\mathbf{K_n}$ averaged over $\mathbf{f}$.

• Second term represents object variability as seen in the mean image.

Thus we define

$$\mathbf{K_g} \equiv \overline{\mathbf{K}}_\mathbf{n} + \mathbf{K}_{\overline{\mathbf{g}}} . \quad (8.347)$$

This division of the overall covariance does not require any assumptions about the form of either $\text{pr}(\mathbf{g}|\mathbf{f})$ or $\text{pr}(\mathbf{f})$.

In particular, it does not require that the noise be object-independent, and it does not require that either the noise or the object be Gaussian.

Doubly stochastic characteristic functions

Suppose the characteristic *functional* of the random object is known:

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \langle \exp[-2\pi i(\mathbf{s}, \mathbf{f})] \rangle \,, \tag{8.94}$$

For signal-independent noise, the corresponding characteristic *function* for the data is

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \psi_{\mathbf{n}}(\boldsymbol{\xi}) \Psi_{\mathbf{f}} \left(\boldsymbol{\mathcal{H}}^{\dagger} \boldsymbol{\xi} \right) \,.$$

For Poisson noise:

$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}[\boldsymbol{\mathcal{H}}^{\dagger} \boldsymbol{\Gamma}(\boldsymbol{\xi})] \,. \tag{8.339}$$

where $\boldsymbol{\Gamma}$ is an operator defined by

$$[\boldsymbol{\Gamma}(\boldsymbol{\xi})]_m = \frac{-1 + \exp(-2\pi i \, \xi_m)}{-2\pi i} \,. \tag{8.338}$$

Ref.: Clarkson *et al.* (2002)

IV. Random objects

Objects in PET and SPECT are random because:

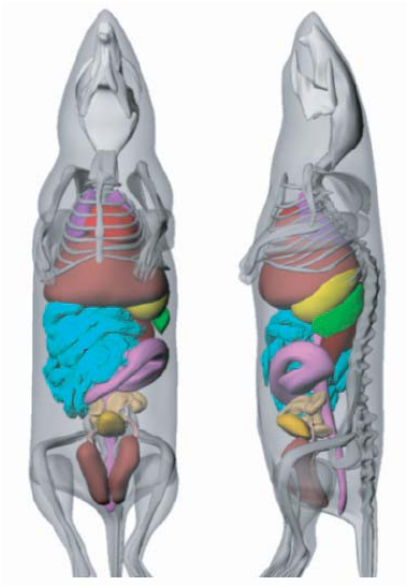
- Organs have random shapes
- Tracer uptake varies randomly from organ to organ
- The fine structure (texture) of uptake within an organ is random

Scatter and attenuation are also random, but we will consider those effects as part of the imaging system.

Some specific object models

- Parametric shape models
- Normal and log-normal texture models
- Stationary and quasistationary models
- Channel and filter-bank models
- Cuspy, kurtotic models
- Lumpy and clustered lumpy models

Random shapes



MOBY mouse phantom developed by Paul Segars

Random shapes are defined by a set of parameters;

PDF on the object $\equiv$ PDF on the shape parameters

Phantom simulation program is way of drawing samples from this PDF

Cuspy, kurtotic densities

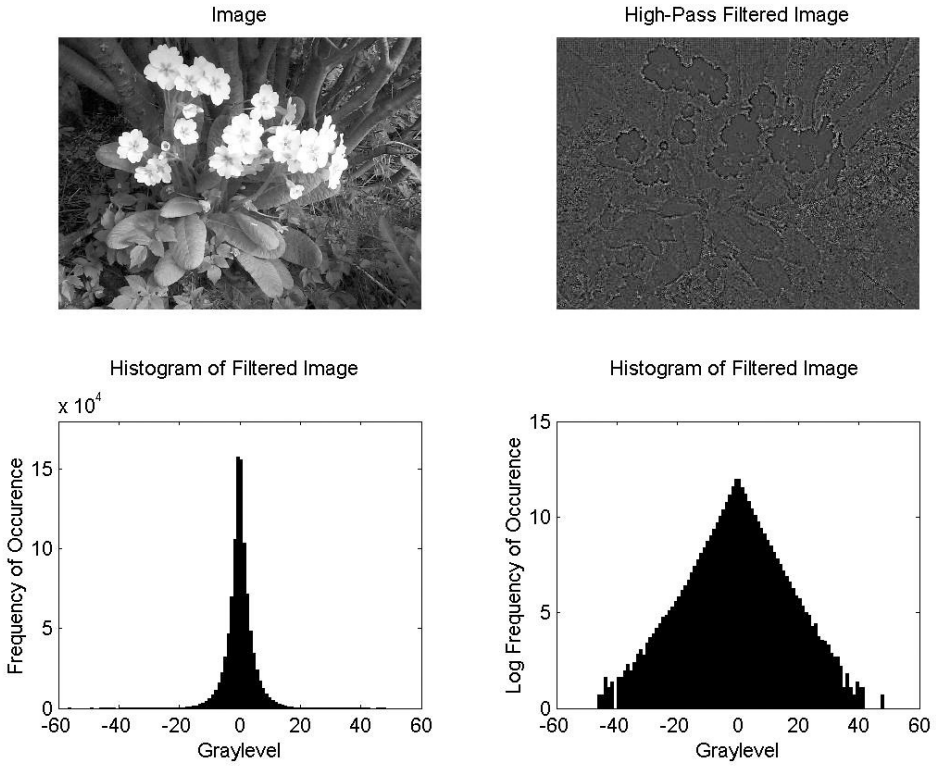


Fig. 8.7 *Top*: A typical image before and after high-pass filtering. *Bottom*: Gray-level histogram of the high-pass filtered image (note that the right plot is vs. log frequency of occurrence).

Wavelet analysis

Wavelet components of an image *may* be:

- Approximately uncorrelated
- Approximately independent
- Cuspy and kurtotic

Filter-bank models for image texture

• Filter-bank methods describe multivariate image statistics in terms of univariate PDFs on the outputs of multiple filters (Mumford et al.)

• Most commonly, the filters are wavelet channels or other bandpass filters

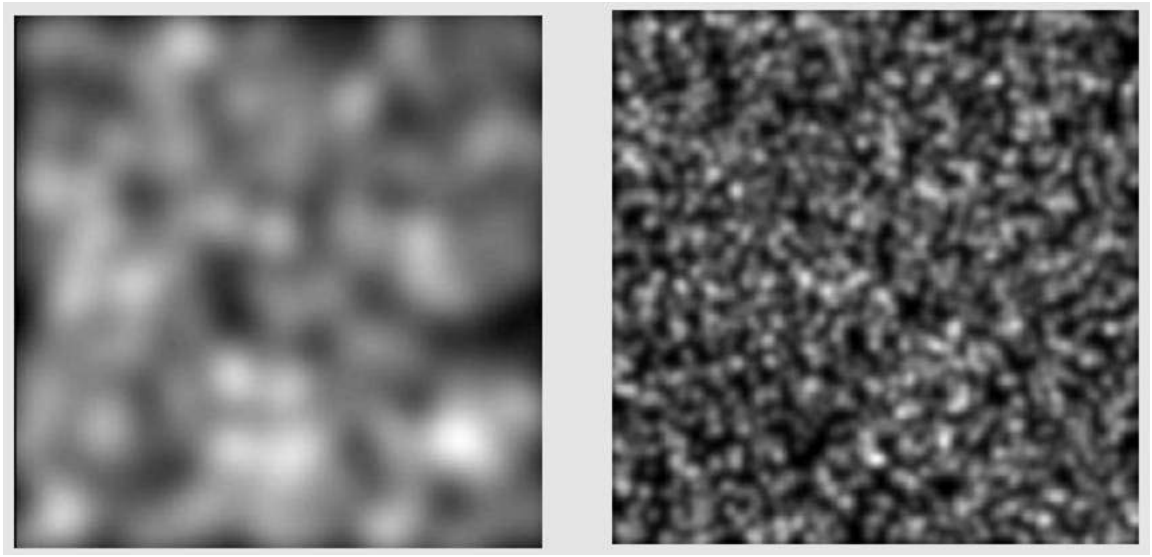
• Premise: Images that have similar histograms on the filter outputs have similar textures (Bergen and Adelson, 1991; Chubb and Landy, 1991)

Methods:

• Heeger and Bergen (1995), Rolland et al. (1998, 2000): Synthesize sample textures from one or more training images

• Zhu et al. (1998): Use maximum entropy to find multivariate PDF consistent with observed univariate PDFs (details in Sec. 8.4.4)

Lumpy backgrounds



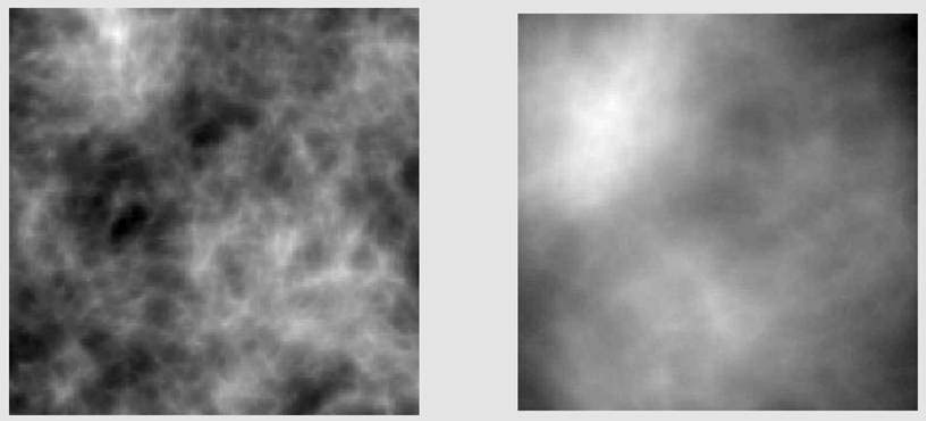
Sample lumpy backgrounds

$$f(\mathbf{r}) = \sum_{n=1}^{N} l(\mathbf{r} - \mathbf{r}_n) \,. \tag{8.303}$$

A common choice for $l(\mathbf{r})$ is a Gaussian spatial distribution

Ref.: Rolland and Barrett, 1992

Clustered lumpy backgrounds



Sample clustered lumpy backgrounds

A cluster of identical blobs forms a *superblob*, and the final model is a superposition of superblobs:

$$f(\mathbf{r}) = \sum_{k=1}^{N_s} \sum_{n=1}^{N_k} l_k(\mathbf{r} - \mathbf{r}_{nk} - \mathbf{R}_k)\,, \tag{8.305}$$

Ref.: Bochud et al., 1999

Statistics of lumpy and clustered lumpy backgrounds

- Models are parametric
- Realistic patterns can be generated by proper choice of parameters
- Analytic expressions for characteristic functionals are known

For lumpy background:

$$\Psi_{\mathbf{f}}(\mathbf{s}) = \exp\left\{-\overline{N} + \int_A d^2 r_n \; b(\mathbf{r}_n) \exp\left[-2\pi i \int_A d^2 r \; s(\mathbf{r})\, l(\mathbf{r}-\mathbf{r}_n)\right]\right\} . \tag{11.152}$$

where $b(\mathbf{r}_n)$ is spatial distribution of lump centers.

All statistical properties of the lumpy background are contained in this characteristic functional.

Estimating parameters in object models

Suppose the object characteristic functional, denoted $\Psi_{\mathbf{f}}(\mathbf{s}|\boldsymbol{\alpha})$ is known except for some parameter set $\boldsymbol{\alpha}$ (e.g., lump density and size)

Recall that the characteristic function of $\mathbf{g}$ is given by:

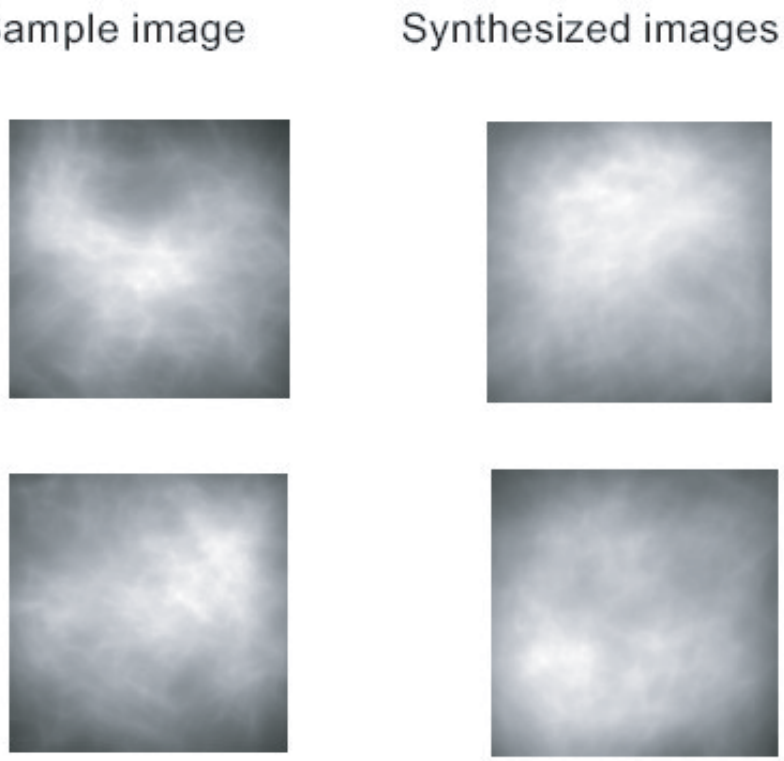
$$\psi_{\mathbf{g}}(\boldsymbol{\xi}) = \Psi_{\mathbf{f}}[\boldsymbol{\mathcal{H}}^{\dagger}\,\boldsymbol{\Gamma}(\boldsymbol{\xi})|\boldsymbol{\alpha}]\,. \tag{8.339}$$

Can estimate $\boldsymbol{\alpha}$ by least-squares fitting of empirical characteristic function (derived from sample images) to $\psi_{\mathbf{g}}(\boldsymbol{\xi})$.

Then generate new images consistent with these statistics.

Ref.: Kupinski et al., 2003c

Estimating parameters in object models – Results



Ref.: Kupinski et al., 2003c

IV. Random systems

PET and SPECT imaging systems are random because:

- System matrix $\mathbf{H}$ may be measured
- Analytic matrix may be based on measured attenuation map
- Scatter may be ignored or only approximately modeled
- Patient may move
- Mechanics of system may not be perfect
- They may adapt to random objects

Triply stochastic dynamic images

Generalize previous doubly stochastic results in two ways:

- Consider dynamic sequence of images
- Allow imaging system to be third source of randomness

Notation:

$$\mathbf{G} = \{\mathbf{g}^{(j)}, j = 1, ..., J\} = \text{ sequence of } J \text{ images}$$

$$\mathbf{F} = \{\mathbf{f}^{(j)}, j = 1, ..., J\} = \text{ sequence of object distributions}$$

$$\mathbf{P} = \{\mathbf{p}^{(j)}, j = 1, ..., J\} = \text{ sequence of system states (PSFs)}$$

Reference:

H. H. Barrett et al.,
Objective Assessment of Image Quality IV (OAIQ IV)
J. Opt. Soc. Am. A. 2006, to be published

Triply stochastic averaging

Consider arbitrary function of the data, $\mathbf{T}(\mathbf{G})$

$$\overline{\overline{\overline{\mathbf{T}(\mathbf{G})}}} = \langle\langle\langle\, \mathbf{T}(\mathbf{G})\,\rangle_{\mathbf{G}|\mathbf{P},\mathbf{F}}\rangle_{\mathbf{P}|\mathbf{F}}\rangle_{\mathbf{F}}$$

$$= \int d\mathbf{F} \int d\mathbf{P} \int d\mathbf{G}\ \mathbf{T}(\mathbf{G})\ \mathrm{pr}(\mathbf{G}|\mathbf{P},\mathbf{F})\ \mathrm{pr}(\mathbf{P}|\mathbf{F})\ \mathrm{pr}(\mathbf{F})\ .$$

$\mathrm{pr}(\mathbf{G}|\mathbf{P},\mathbf{F}) \Rightarrow$ PDF of measurement noise for given system and object

$\mathrm{pr}(\mathbf{P}|\mathbf{F}) \Rightarrow$ PDF of system PSF for given object sequence

$\mathrm{pr}(\mathbf{F}) \Rightarrow$ PDF of sequence of random objects

Covariance decomposition

$$\mathbf{K_G} = \overline{\overline{\mathbf{K}}}_{\mathbf{G}}^{noise} + \overline{\mathbf{K}}_{\overline{\mathbf{G}}}^{PSF} + \mathbf{K}_{\overline{\overline{\mathbf{G}}}}^{obj} \, ,$$

where

$$\overline{\overline{\mathbf{K}}}_{\mathbf{G}}^{noise} \equiv \langle \langle \langle [\mathbf{G} - \overline{\mathbf{G}}\,] \, [\mathbf{G} - \overline{\mathbf{G}}\,]^t \rangle_{\mathbf{G}|\mathbf{P},\mathbf{F}} \rangle_{\mathbf{P}|\mathbf{F}} \rangle_{\mathbf{F}} \; ;$$

$$\overline{\mathbf{K}}_{\overline{\mathbf{G}}}^{PSF} \equiv \langle \langle [\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\,] \, [\overline{\mathbf{G}} - \overline{\overline{\mathbf{G}}}\,]^t \rangle_{\mathbf{P}|\mathbf{F}} \rangle_{\mathbf{F}} \; ;$$

$$\mathbf{K}_{\overline{\overline{\mathbf{G}}}}^{obj} \equiv \langle [\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}\,] \, [\overline{\overline{\mathbf{G}}} - \overline{\overline{\overline{\mathbf{G}}}}\,]^t \rangle_{\mathbf{F}} \; .$$

Thus the overall covariance matrix for a triply stochastic image sequence can be rigorously decomposed into three terms representing, respectively, the contributions from measurement noise, from the random PSF and from randomness in the object being imaged.

For many more details, see OAIQ IV

VI. Challenges, unsolved problems, areas for future research

Image statistics

- Statistics of nonlinear reconstructions with random objects
- More analysis of likelihood for estimation problems
- Spatiotemporal statistics for dynamic studies
- Statistics of adaptive image sequences
- More analysis of mixed Poisson and Gaussian noise

Object statistics

- Characteristic functionals for random shapes
- Empirical validation of analytic models
- Texture databases
- Empirical organ-uptake statistics
- Fast simulation methods for sample covariance matrices

Random systems

- Characteristic functionals for general random systems
  (Already done for atmospheric turbulence!)

- Analysis of effect of noisy, measured $\mathbf{H}$.

- Analysis of effect of measured attenuation

- Statistics of patient motion

- Empirical statistics for scatter

- More analysis of adaptive, time-varying systems

- Sample and simulation methods for finding the system covariance