BME I5100: Biomedical Signal Processing

Random Variables

Lucas C. Parra
Biomedical Engineering Department
City College of New York

CCNY

Schedule

Week 1: Introduction
Linear, stationary, normal - the stuff biology is **not** made of.

Week 1-4: Linear systems
Impulse response
Moving Average and Auto Regressive filters
Convolution
Discrete Fourier transform and z-transform
Sampling

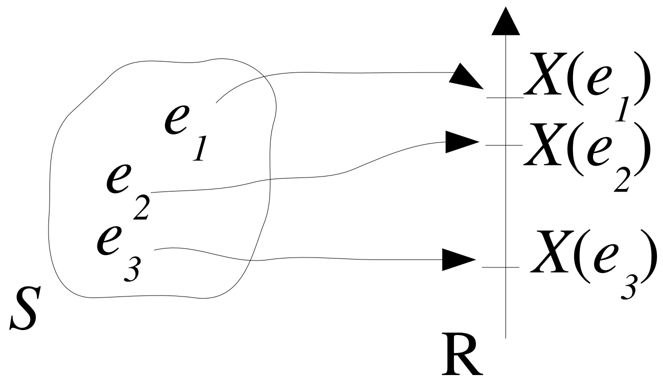
Week 5-8: Random variables and stochastic processes
Random variables
Moments and Cumulants
Multivariate distributions
Statistical independence and stochastic processes

Week 9-14: Examples of biomedical signal processing
Probabilistic estimation
Linear discriminants - **detection** of motor activity from MEG
Harmonic analysis - **estimation** of hart rate in ECG
Auto-regressive model - **estimation** of the spectrum of 'thoughts' in EEG
Matched and Wiener filter - **filtering** in ultrasound
Independent components analysis - **analysis** of MEG signals

Random Variables

Probabilities are defined for discrete events, e.g. outcome of a coin flip.

A **random variable** X is a **map** from an event $e \in S$ to an observed real value $X(e)$.

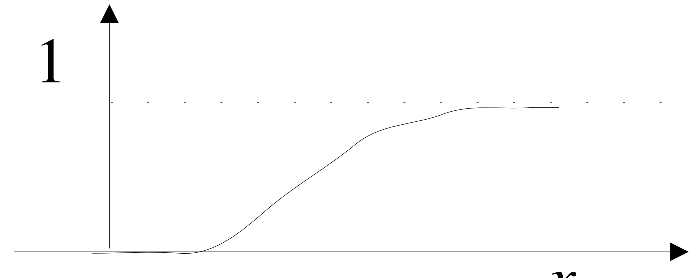


The event can be thought of as a specific realization of a random system. $X(e)$ is an associate observation, e.g. Patients body the temperature measured ad 6AM.

Random Variables - Probability density

Probability distribution defined a the probability that X

F(x) = “Probability of findin value smaller than x”

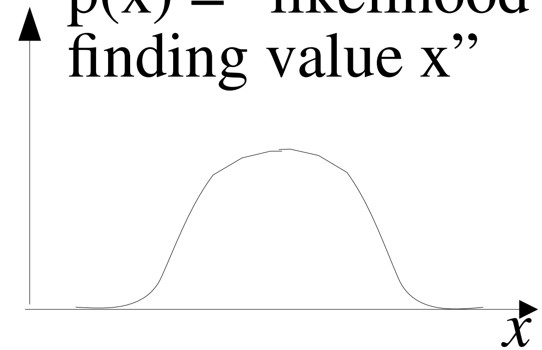


$$F_X(x) = Pr(X \leqslant x)$$

$$F_X(-\infty) = 0, \quad F_X(\infty) = 1$$

Probability density function (**PDF**)

p(x) = “likelihood of finding value x”



$$p_X(x) = \frac{\partial}{\partial x} F_X(x)$$

Alternatively, a **PDF** is a real valued function with

$$\int_{-\infty}^{\infty} dx\, p_X(x) = 1, \quad p_X(x) \geqslant 0$$

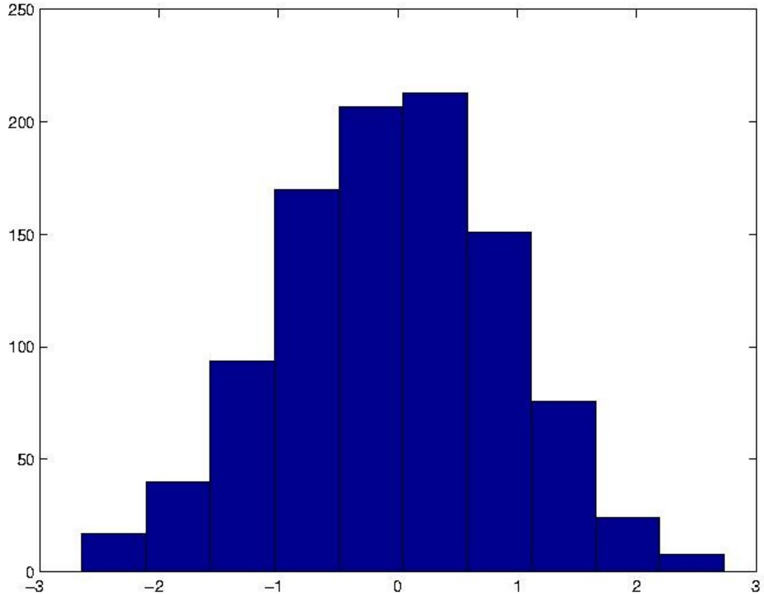
Random Variables - Histogram

Estimate of probability density is the **histogram**

$$p_X(x) \approx \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \propto Pr(x \leqslant X \leqslant x + \Delta x)$$

Where we measure the the likelihood $Pr(x \leqslant X \leqslant x + \Delta x)$ by counting how many samples fall within x and $x + \Delta x$.

```
>> hist(randn(1000,1))
```



Another way of assessing the structure of a pdf are *moments* and *cumulants*.

Random Variables -Moments

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

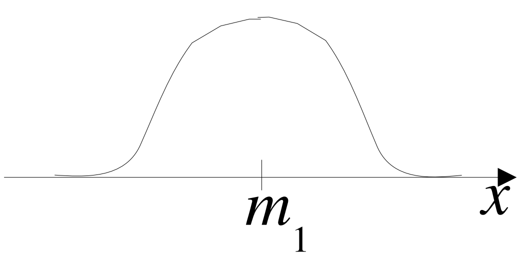
$$E[f(X)] = \int_{-\infty}^{\infty} dx\, p(x) f(x)$$

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**

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

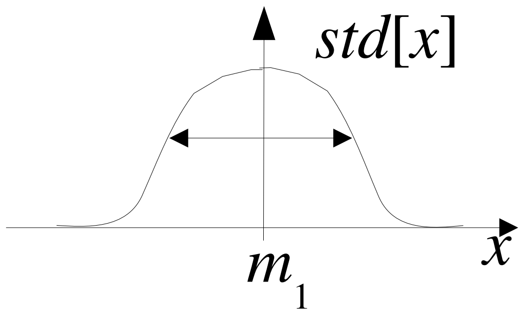
Random Variables -Moments and Cumulants

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**

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



RV -Moment generating function

Consider the Laplace transform of the PDF (up to sign of *t*)

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

Given $E[e^{tx}]$ the PDF is fully determined by the inverse Laplace transform. Consider now the Taylor expansion

$$E[e^{tx}] = \sum_{n=0}^{\infty} \frac{m_n}{n!} t^n$$

Note that the expansion coefficients are the moments of the PDF:

$$m_n = \frac{\partial^n}{\partial t^n} E[e^{tx}]_{t=0} = E[x^n]$$

We call $E[e^{tx}]$ therefore the **moment generating function.**
Given all moments $E[e^{tx}]$ is fully determined and so is the PDF.

RV -Cumulants

Cumulants c_n of order n are defined as the expansion coefficients of the logarithm of the moment generating function:

$$\ln E[e^{tx}] = \sum_{n=0}^{\infty} \frac{c_n}{n!} t^n \qquad c_n = \frac{\partial^n}{\partial t^n} \ln E[e^{tx}]_{t=0}$$

Cumulants are a convenient way of describing properties of the PDF:

c_1 = mean

c_2 = variance, measures spread[2] around the mean.

$c_3 \propto$ skew, measures asymmetry around mode (=0 for symmetric PDF)

$c_4 \propto$ kurtosis, measures length of tails (=0 for Gaussians)

$$c_1 = \mu$$

$$c_2 = \sigma^2$$

$$c_3 = E[(x-\mu)^3]/\sigma^3$$

$$c_4 = E[(x-\mu)^4]/\sigma^4$$

Assignment 6: Show that mean and variance are first two cumulants.

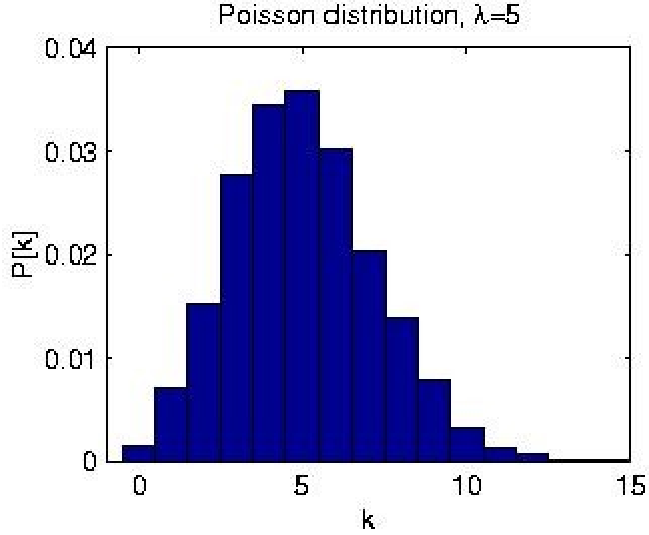
Random Variables -Poisson distribution

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

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: $E[k] = \lambda$

Second moment: $E[k^2] = \lambda^2 + \lambda$

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

Fano factor: $F = var[k] / E[k] = 1$

Fano-factor is a easy metric to assert that a count is **not** Poisson. $F \neq 1$ often used to establish that spike counts are more "interesting" than Poisson.

Random Variables -Exponential distribution

Exponential random variable has a pdf:

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

Mean: $E[x]=\lambda^{-1}$

Variance: $var[x]=\lambda^{-2}$

Spike train with constant firing rate λ (number of spikes per unit time) for which the occurrence of a spike is independent of previous spikes has exponentially distributed inter spike intervals (ISI).

Assignment 7:
Generate Poisson distributed samples and measure F.
Could the counts in the two spike trains in `spike.mat` be Poisson?

Sampling from a continuous distribution

Goal: Generate random numbers y following a desired pdf $p(y)$.

Approach: Draw x from an uniform distribution $p(x) = const$ and apply a nonlinearity $y = f(x)$.

$$p(x) = \left|\frac{dy}{dx}\right| p(y) = const.$$

$$x = \int_0^x dx' p(x') = \int_{-\infty}^{y} p(y') dy' \equiv F_Y(y)$$

Result: The desired non-linear transformation is the inverse cumulative density, or inverse of the distribution function:

$$y = f(x) = F_Y^{-1}(y)$$

Example: Draw N samples from exponential distribution with mean m:

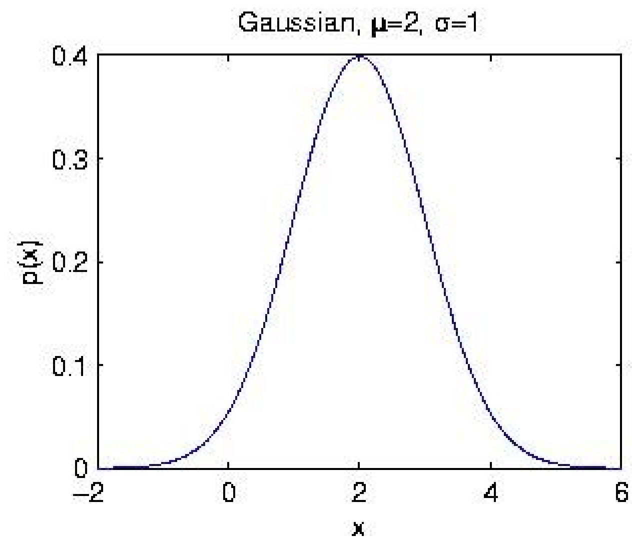
```
dy=m/1000; y=0:dy:10*m;
p = exp(-y/m)/m;
F = cumsum(p)*dy;
yrand = interp1(F,y,rand(N,1));
```

Note: This simple technique may not do a good job with the tails.

Random Variables -Normal distribution

Perhaps the most important distribution is the normal distribution with Gaussian PDF:

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Often written in short as $N(\mu,\sigma^2)$.

This defines a density because it is positive and normalized

$$\int_{-\infty}^{\infty} dx \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)=1$$

Which is easy to show using $$\int_{-\infty}^{\infty} dx\, e^{-x^2}=\sqrt{\pi}$$

Random Variables -Normal distribution

The moments can be computed using the moment generating function

$$E[e^{tx}] = \int_{-\infty}^{\infty} dx\, e^{tx}\, p(x) = (\sigma\sqrt{2\pi})^{-1} \int_{-\infty}^{\infty} dx\, e^{tx}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= e^{\mu t + \sigma^2 t^2/2}$$

Mean:

$$\left(\frac{d\,E[e^{tx}]}{dt}\right)_{t=0} = \mu$$

Second moment:

$$\left(\frac{d^2 E[e^{tx}]}{dt^2}\right)_{t=0} = \sigma^2 + \mu^2$$

Variance:

$$var[x] = \sigma^2$$

To estimate best fitting Gaussian simply measure mean and variance!

Note that all higher cumulants are zero:

$$\ln E[e^{tx}] = \mu t + \sigma^2 t^2/2$$

Random Variables -Normal distribution

Product of Gaussians is a Gaussian

$$e^{-\frac{x^2}{2\sigma_1^2}} e^{-\frac{x^2}{2\sigma_2^2}} = e^{-\frac{x^2}{\sigma_3^2}} \qquad \sigma_3 = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

Convolution of Gaussians is a Gaussian

$$\int_{-\infty}^{\infty} dx\, e^{-\frac{x^2}{2\sigma_1^2}} e^{-\frac{(y-x)^2}{2\sigma_2^2}} \propto e^{-\frac{y^2}{2\sigma_3^2}} \qquad \sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}$$

Random Variables -Sample average

Let X_1, X_2, ... be independently an identically drawn samples from an arbitrary distribution with mean μ and variance σ^2.

Consider the **sample average**:

$$W_n = \frac{1}{n} \sum_{k=1}^{n} X_k$$

The **Law of Large Numbers** states that sample average converges to the ensemble average

$$\lim_{n \to \infty} W_n = E[X_k]$$

The **Central Limit Theorem** states that sample average is normal

$$\lim_{n \to \infty} p(W_n) = N(\mu, \sigma^2/n)$$