Probability Review

(many slides from Octavia Camps)

Intuitive Development

- Intuitively, the probability of an event **a** could be defined as:

$$P(a) = \lim_{n \to \infty} \frac{N(a)}{n}$$

Where N(a) is the number that event a happens in n trials

More Formal:

- Ω is the Sample Space:
  - Contains all possible outcomes of an experiment
- $\omega \in \Omega$ is a single outcome
- $A \in \Omega$ is a set of outcomes of interest

1. $P(A) \geq 0 \forall A \in \Omega$
2. $P(\Omega) = 1$
3. $A_i \cap A_j = \emptyset \forall i, j \Rightarrow P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
4. $P(\emptyset) = 0$

Independence

- The probability of independent events A, B and C is given by:

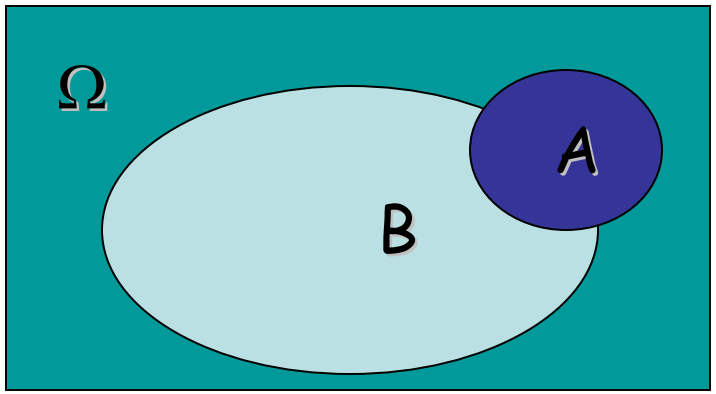
$$P(ABC) = P(A)P(B)P(C)$$

A and B are independent, if knowing that A has happened does not say anything about B happening

Conditional Probability

- One of the most useful concepts!

$$P(A|B) = \frac{P(AB)}{P(B)}$$



Bayes Theorem

- Provides a way to convert *a-priori* probabilities to *a-posteriori* probabilities:

$$P(A|B)P(B) = P(B|A)P(A)$$

Probabilities sum to 1

- All facts are either true or false

  $p(d) + p(\sim d) = 1$

- Marginalization

  $p(x|y)p(y) + p(x|\sim y)p(\sim y) = p(x)$

  $p(x) = p(x|y=1)p(y=1) + p(x|y=2)p(y=2)$
  $+ p(x|y=3)p(y=3) \ldots$
  $= \Sigma\, p(x|y=i)p(y=i)$

Bayes Rule

$$p(h|e) = (p(e|h) / p(e)) * p(h)$$

posterior prior

Prior: **initial estimate of probability of hypothesis**

Posterior: **estimate of probability after seeing evidence**

Proof: follows directly from definition

$p(h|e) = p(h,e) / p(e)$

What does probability mean?

Frequentists: Probability as expected frequency

- *P*(“*heads*”) = 0.5 ~ “If we flip 100 times, we expect to see about 50 heads.”

Subjectivists: Probability as degree of belief

- *P*(“*heads*”) = 0.5 ~ “On the next flip, it’s an even bet whether it comes up heads or tails.”
- *P*(“*rain tomorrow*”) = 0.8
- *P*(“*bin Laden is dead*”) = 0.1
- . . .

What does probability mean?

Frequentists: Probability as expected frequency

- $P(A) = 1$: A will always occur.
- $P(A) = 0$: A will never occur.
- $0.5 < P(A) < 1$: A will occur more often than not.

Subjectivists: Probability as degree of belief

- $P(A) = 1$: believe A is true.
- $P(A) = 0$: believe A is false.
- $0.5 < P(A) < 1$: believe A is more likely to be true than false.

Bayesian inference

- Definition of conditional probability:

$$P(A, B) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

- Bayes' rule:

$$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(D)}$$

- "Posterior probability": $P(H \mid D)$
- "Prior probability": $P(H)$
- "Likelihood": $P(D \mid H)$

Bayesian inference

- Bayes' rule: $$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(D)}$$
- What makes a good scientific argument? $P(H|D)$ is high if:
  - Hypothesis is plausible: $P(H)$ is high
  - Hypothesis strongly predicts the observed data:

    $P(D|H)$ is high
  - Data are surprising: $P(D)$ is low

Bayesian Inference

- Predict future observations by marginalizing over hypotheses

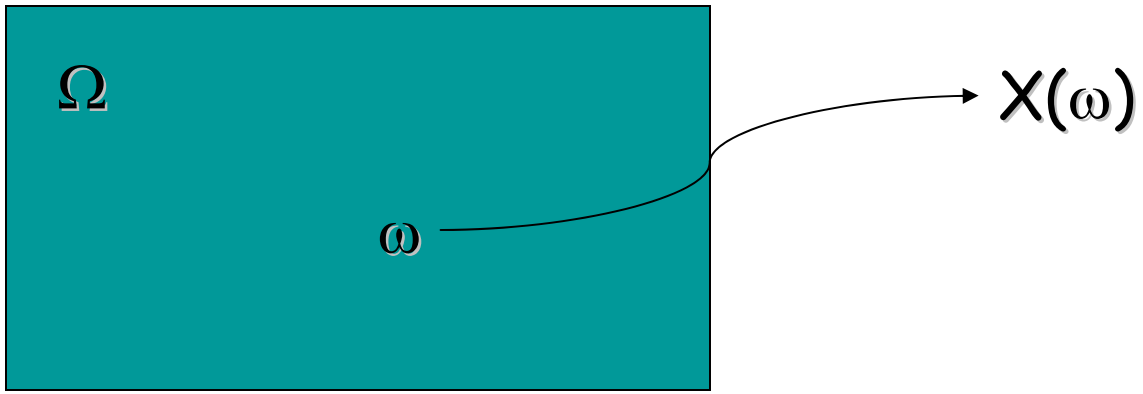
$$P(d_1) = \Sigma_h \, p(d_1 \,|h) p(h|D)$$

$$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(D)}$$

Many plausible hypotheses -> similarity
One plausible hypothesis -> rule

Random Variables

- A (scalar) random variable X is a function that maps the outcome of a random event into real scalar values



Random Variables Distributions

- Cumulative Probability Distribution (CDF):

$$F_X(x) = P(X \leq x)$$

- Probability Density Function (PDF):

$$p_X(x) = \frac{dF_X(x)}{dx}$$

Random Distributions:

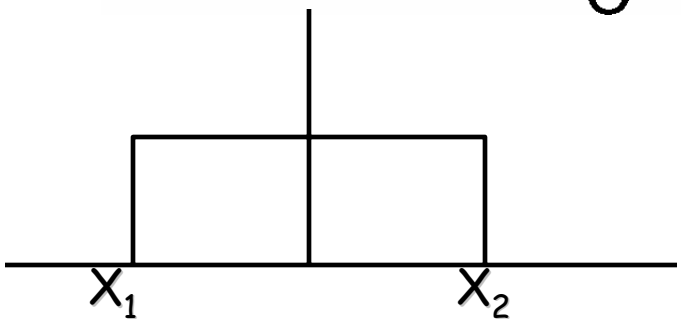
- From the two previous equations:

$$\int_{-\infty}^{\infty} p_X(x)dx = 1.0$$

Uniform Distribution

- A R.V. X that is uniformly distributed between x_1 and x_2 has density function:

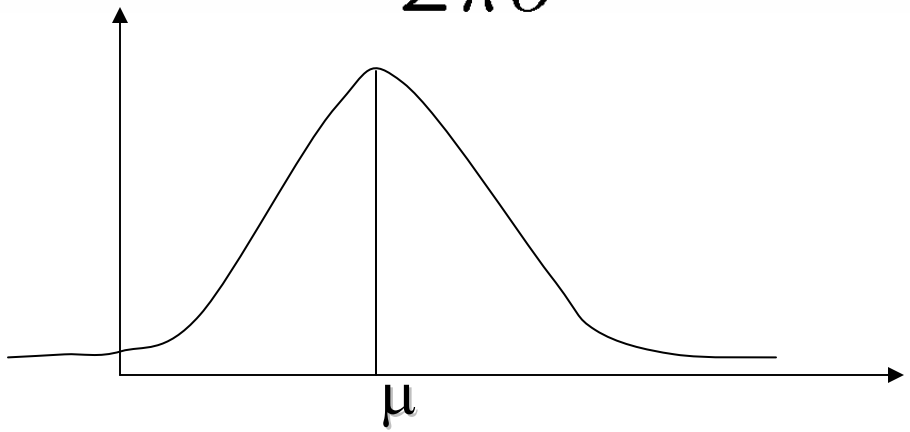
$$p_X(x) = \begin{cases} \frac{1}{x_2 - x_1} & x_1 \leq x \leq x_2 \\ 0 & otherwise \end{cases}$$



Gaussian (Normal) Distribution

- A R.V. X that is normally distributed has density function:

$$p_X(x) = \frac{1}{2\pi\sigma} \exp -\frac{(x-\mu)^2}{2\sigma^2}$$



Statistical Characterizations

- Expectation (Mean Value, First Moment):

$$E(X) = \int_{-\infty}^{\infty} x p_X(x) dx$$

- **Second Moment:**

$$E(X^2) = \int_{-\infty}^{\infty} x^2 p_X(x) dx$$

Statistical Characterizations

- Variance of X:

$$\begin{aligned} Var(X) &= E\{[X - E(X)]^2\} \\ &= \int_{-\infty}^{\infty} (x - E[X])^2 p_X(x) dx \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

- Standard Deviation of X:

$$\sigma_X = \sqrt{Var(X)}$$

Mean Estimation from Samples

- Given a set of N samples from a distribution, we can estimate the mean of the distribution by:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Variance Estimation from Samples

- Given a set of N samples from a distribution, we can estimate the variance of the distribution by:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2$$

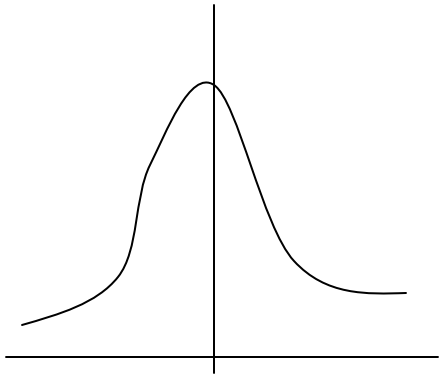
Image Noise Model

- Additive noise:
  - Most commonly used

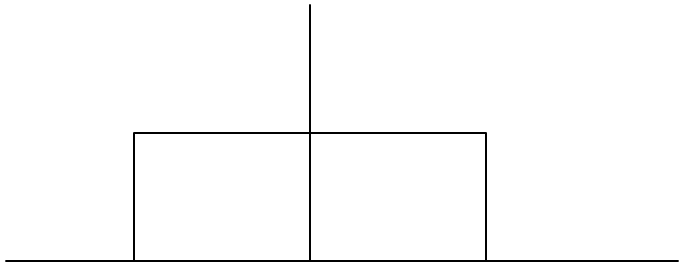
$$\hat{I}(i, j) = I(i, j) + N(i, j)$$

Additive Noise Models

- Gaussian
  - Usually, zero-mean, uncorrelated



- Uniform



Measuring Noise

- Noise Amount: SNR = σ_s / σ_n

- Noise Estimation:
  - Given a sequence of images $I_0, I_1, \ldots I_{N-1}$

$$\bar{I}(i,j) = \frac{1}{N}\sum_{k=0}^{N-1} I_k(i,j)$$

$$\sigma(i,j) = \sqrt{\frac{1}{N-1}\sum_{k=0}^{N-1}(\bar{I}(i,j) - I_k(i,j))^2}$$

$$\sigma_n = \frac{1}{RC}\sum_{i=0}^{R-1}\sum_{j=0}^{C-1}\sigma(i,j)$$

Good estimators

Data values z are random variables
A parameter θ describes the distribution
We have an estimator $\varphi(z)$ of the unknown parameter θ.

If

$$E(\varphi(z) - \theta) = 0 \quad \text{or}$$

$$E(\varphi(z)) = E(\theta)$$ the estimator $\varphi(z)$ is unbiased

Least Squares (LS)

$$Au = b.$$

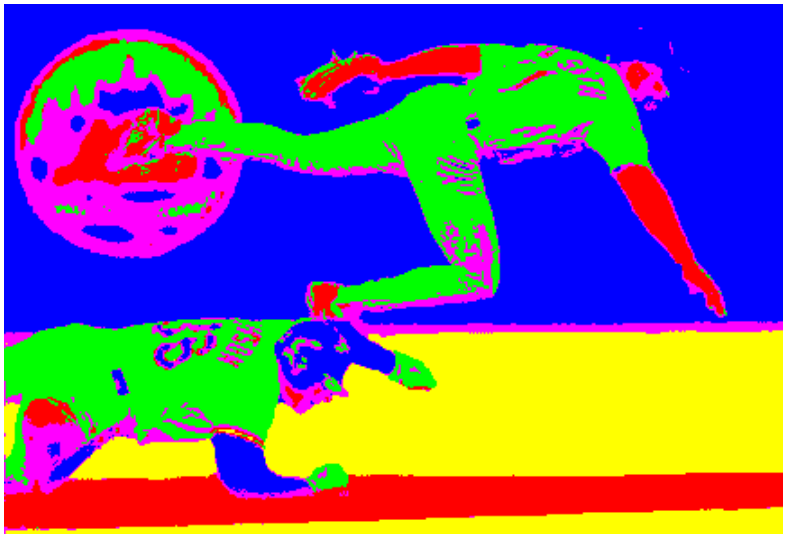
If errors only in b

$$Au = b + \delta b.$$

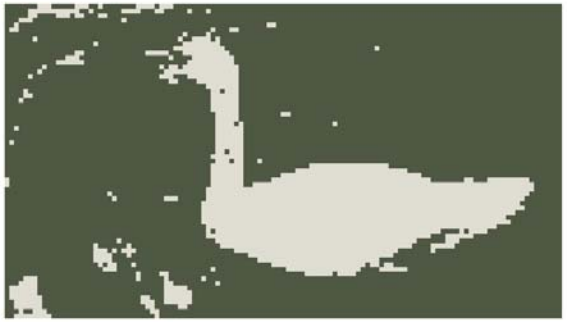
Then LS is unbiased

$$u_l = (A^t A)^{-1} A^t b.$$

Clustering pixels: segmentation and recognition



Clustering Color/Intensity



Group together pixels of similar color/intensity.

Agglomerative Clustering

- Cluster = connected pixels with similar color.
- Optimal decomposition may be hard.
  - For example, find k connected components of image with least color variation.
- Greedy algorithm to make this fast.

Clustering Algorithm

- Initialize: Each pixel is a region with color of that pixel and neighbors = neighboring pixels.
- Loop
  - Find adjacent two regions with most similar color.
  - Merge to form new region with:
    - all pixels of these regions
    - average color of these regions.
    - All neighbors of either region.
  - Stopping condition:
    - No regions similar
    - Find *k* regions.

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

23	25	19	21	23
18	22	24.5	24.5	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

23	25	19	21	23
18	22	24.33	24.33	24.33
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

23	25	19	21	23
18	22	24.33	24.33	24.33
19.5	19.5	26	28	22
3	3	7	8	26
1	3	5	4	24

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

23	25	19	21	23
18	22	24.33	24.33	24.33
19.5	19.5	26	28	22
3	3	7.5	7.5	26
1	3	5	4	24

Example

. . .

Example

23	25	19	21	23
18	22	24	25	24
20	19	26	28	22
3	3	7	8	26
1	3	5	4	24

22.9	22.9	22.9	22.9	22.9
22.9	22.9	22.9	22.9	22.9
22.9	22.9	22.9	22.9	22.9
4.25	4.25	4.25	4.25	22.9
4.25	4.25	4.25	4.25	22.9

Clustering complexity

- n pixels. (not here n referes to total number of pixels)
- Initializing:
  - O(n) time to compute regions.
- Loop:
  - O(n) time to find closest neighbors (could speed up).
  - O(n) time to update distance to all neighbors.
- At most n times through loop so O(n*n) time total.

Agglomerative Clustering: Discussion

- Start with definition of good clusters.
- Simple initialization.
- Greedy: take steps that seem to most improve clustering.
- This is a very general, reasonable strategy.
- Can be applied to almost any problem.
- But, not guaranteed to produce good quality answer.

Parametric Clustering

- Each cluster has a mean color/intensity, and a radius of possible colors.
- For intensity, this is just dividing histogram into regions.
- For color, like grouping 3D points into spheres.