Basic Outline of Midterm Topics

March 7, 2019

Basic Probability Topics

- Random Variables
- **Expected Values**
- **Conditional Probability and Bayes Theorem**
- Transformation of Random Variables

Gaussian PDFs

- Univariate: $w[n] \sim N(\mu, \sigma^2)$
  - $f[n, \theta] + w[n] \sim N(\mu + f[n, \theta], \sigma^2)$
  - $f[n, \theta] \cdot w[n] \sim N(\mu, \sigma^2 f[n, \theta]^2)$
  - $p(w[n]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-1}{2\sigma^2}(w[n] - \mu)^2)$
- Multivariate: $\mathbf{w} \sim N(\mu, \mathbf{C})$
  - $\mathbf{C}$ is an $N \times N$ positive semidefinite matrix.
  - $p(w[n]) = \frac{1}{\sqrt{(2\pi)^N |C|}} \exp(\frac{-1}{2}(\mathbf{w} - \mu)^T \mathbf{C}^{-1}(\mathbf{w} - \mu))$
- You can't divide by $\mathbf{C}$ in the multivariate distribution.

Models

- Usually we have a set of samples, observations, measurements, which we call $\mathbf{x} = [x[0], ..., x[N-1]]$.
- A Model is a function that tries to describe $\mathbf{x}$ using parameters θ and noise $w[n]$, e.g.:

$$x[n] = f(n, \theta, w[n])$$

- In this class, we pretty much deal with additive noise, i.e.

$$x[n] = f(n, \theta) + w[n]$$

- An estimator, $\hat{\theta}(\mathbf{x})$, tries to "solve" this equation for θ.
- Because of $w[n]$, $x[n]$ is a random variable, so $\hat{\theta}$ is also a random variable.

Performance

- Bias: $\mathbb{E}[\hat{\theta} - \theta]$
- Variance: $\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$
- Mean Squared Error: $\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathrm{var}(\hat{\theta}) + \mathbf{bias}(\hat{\theta})^2$

Sufficient Statistics

- Sufficient Statistic - building block of estimators, contain all information necessary to estimate $\hat{\theta}$.
- Factorization Theorem
- Exponential Family of Distributions

Cramer-Rao Bound

- Fisher Information Matrix - Several Formulas to find this value.
- CRB - scalar and vector cases.
- CRB is used for unbiased estimators, general information inequality used for bias - but bias is not always easy to compute so this formulation isn't used much.
- Cauchy-Schwartz- Useful Inequality
- Efficient Estimator: $\mathrm{var}(\hat{\theta}) = CRB(\theta)$

Linear Models

- General formulation:

$$x[n] = \mathbf{h_i}^T \theta + w[n] \implies \mathbf{x} = \mathbf{H}\theta + \mathbf{w}$$

- The least squares solution is the minimum variance efficient estimator.
- We know the formula for the estimator, and for the variance/CRB of the estimator.
- What to do for colored noise? $C \neq \sigma^2 \mathbb{I}$.
- When $\mathbf{w}$ is not gaussian, we have the Best Linear Unbiased Estimator.

Maximum Likelihood

- Given $x[n] = f(n, \theta) + w[n]$, where the pdf of $w[n]$ is known, find the pdf:

$$p(x[n]; \theta)$$

- For this class, we assume that all of the $x[n]'s$ are independent, so that

$$p(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} p(x[n]; \theta)$$

- Maximum Likelihood Estimation:

$$\hat{\theta}(\mathbf{x}) = \arg\max_{\theta} p(\mathbf{x}; \theta) = \arg\max_{\theta} \log(p(\mathbf{x}; \theta)) = \arg\max_{\theta} \sum_{n=0}^{N-1} \log(p(x[n]; \theta))$$

- To maximize, set gradient/derivative equal to 0, check the second derivative if there is more than one 0.

Maximum Likelihood Continued

- Maximum Likelihood Estimators are asymptotically unbiased, i.e. as $N \to \infty$ the bias is 0.
- They are asymptotically efficient - i.e. as $N \to \infty$ they hit the CRB.
- Only pdfs with efficient estimators belong to the exponential family.

Bayesian Estimation

- Bayes Theorem: $p(\theta|x) = \frac{p(x,\theta)}{p(x)} \propto p(x,\theta) = p(x|\theta)\pi(\theta)$
- We want the posterior $p(\theta|x)$.
- Types of priors: conjugate, proper, improper, Jeffrey's, informative, non-informative.
- Sequential: $p(\theta|x_1, x_2) \propto p(x_2|\theta, x_1)p(\theta|x_1) \propto p(x_2|\theta, x_1)p(x_1|\theta)\pi(\theta)$
- Posterior Predictive distribution: $p(x_2|x_1)$
- Bayesian Sufficient Statistics - have their version of the factorization theorem.