Learning Objectives for STA 360/601

(What you should know for the final.)

1 General concepts

(References: Chapter 1 (Foundations), Hoff chapter 1 and 2.2–2.6, mathematicalmonk PP 2.1–5.5, ML 7.1–7.6)

- Know the difference between probability and statistics.
- Bayes' theorem: Know the formula for Bayes' theorem, understand what it means, and know how to apply it.
- Be able to explain the Bayesian approach to statistics.
- Proportionality: Know the definition, know how to use it to derive conditional distributions such as the posterior, and understand why it works (if two distributions are proportional, they are equal).
- Cast of characters: Know the definitions (mathematical formulas) of, and understand the purpose of:
  - likelihood / generating distribution
  - prior
  - posterior
  - marginal likelihood
  - posterior predictive
  - loss function
  - posterior expected loss
  - risk / frequent risk
  - integrated risk

2 Probability distributions

(References: Hoff page 253) Note: You are not expected to memorize the form of each distribution (e.g., the p.d.f./p.m.f. and c.d.f.). A sheet with common distributions will be provided on the exam.

- Understand when a given probability distribution would be appropriate for modeling a given type of data set (e.g., discrete versus continuous, appropriate range of values)

- Distributions to be familiar with:
  - Discrete: Geometric, Bernoulli, Binomial, Poisson, Uniform
  - Continuous (univariate): Exponential, Uniform, Gamma, Beta, Pareto, Normal, Inverse Gamma, Cauchy, t-distribution
  - Continuous (multivariate): Multivariate Normal/Gaussian, Wishart, Inverse Wishart
- Be able to give examples of when a distribution would be appropriate/inappropriate.

3 Analytical derivations

(References: Chapters 1, 3, 4, 6, Hoff chapters 3, 5, 6, 7)

- Be able to derive:
  - the posterior,
  - the marginal likelihood, and
  - the posterior predictive,

  for simple models with conjugate priors, such as:
  - Bernoulli(θ) with Beta($\theta|a, b$) prior
  - Binomial(n, θ) with Beta($\theta|a, b$) prior
  - Geometric(θ) with Beta($\theta|a, b$) prior
  - Uniform($0, \theta$) with Pareto($\theta|\alpha, c$) prior
  - Exp(θ) with Gamma($\theta|a, b$) prior
  - Gamma(α, θ) with Gamma($\theta|a, b$) prior
  - Poisson(θ) with Gamma($\theta|a, b$) prior
  - $\mathcal{N}(\mu, \lambda^{-1})$ with $\mathcal{N}(\mu|\mu_0, \lambda_0^{-1})$ prior
  - $\mathcal{N}(\mu, \lambda^{-1})$ with Gamma($\lambda|a, b$) prior
  - $\mathcal{N}(\mu, \Lambda^{-1})$ (multivariate) with $\mathcal{N}(\mu|\mu_0, \Lambda_0^{-1})$ prior
  - $\mathcal{N}(\mu, \Lambda^{-1})$ (multivariate) with Wishart($\Lambda|S, \nu_0$) prior

4 Decision theory

(References: Chapter 1 (Foundations), mathematicalmonk ML 3.1–3.4 and 11.1–11.8)

- Understand the decision theoretic setup: state, observation, action, loss.
- Know the Bayesian approach to decision theory (minimize posterior expected loss).
- Know the definition of a Bayes procedure.

- Be able to compute the Bayes procedure (in closed-form when possible) for:
  - 0–1 loss
  - square loss
  - other simple loss functions, such as quadratic functions of the state and action
- Be able to give examples of real-world decision problems that could be addressed using decision theory.
- Know the definition of admissibility.

5 Exponential families

(References: Chapter 3 (Exp Fams and Conj Priors), Hoff section 3.3, mathematicalmonk ML 5.1–5.4)

- Know the definition of a one-parameter exponential family.
- Know the definition of a (multi-parameter) exponential family.
- Be able to show that a given collection of distributions is an exponential family (one-parameter or multi-parameter).
- Be able to identify the sufficient statistics function for a given exponential family (one-parameter or multi-parameter).
- Know the definition of natural form / canonical form, and be able to put a given exponential family into natural form.
- Be able to give examples of exponential families.

6 Conjugate priors

(References: Chapter 3 (Exp Fams and Conj Priors), Hoff section 3.3, mathematicalmonk ML 7.4)

- Know the definition of a conjugate prior family.
- Be able to show that a given collection of distributions is a conjugate prior for a given generator/likelihood family.
- Be able to show that mixtures of conjugate priors are conjugate priors.
- Know how to construct a conjugate prior for an exponential family.
- Understand that conjugate priors are not unique.
- Be able to give examples of conjugate priors.

7 Univariate Normal model

(References: Chapter 4 (Univariate Normal model), Hoff chapter 5, mathematicalmonk ML 7.9–7.10)

- Understand when the Normal model is appropriate.
- Know the basic properties of the normal distribution: mean, median, mode; symmetric; 95% probability inside ± 1.96 sigma.
- Know the relationship between the standard deviation, variance, and precision.
- Know the formula for the distribution of linear combinations of independent normals.
- Know how to construct a conjugate prior for the mean (Normal–Normal model).
- Know how to construct a conjugate prior for the mean and precision (NormalGamma–Normal model)
- Be able to derive the posterior for the Normal–Normal model. (You are not expected to memorize it.)
- Be able to derive the posterior for the NormalGamma–Normal model. (You are not expected to memorize it.)
- Be able to choose appropriate values for the prior parameters (hyperparameters).
- Understand the relationship between the Gamma distribution and Inverse Gamma distribution, and understand how they are used for constructing priors on the precision and variance, respectively.
- Know that the Normal model is sensitive to outliers.

8 Monte Carlo

(References: Chapter 5 (Monte Carlo approx), Hoff chapter 4, mathematicalmonk ML 17.1–17.4)

- Know what a (simple) Monte Carlo approximation is.
- Know what kinds of things can be approximated using Monte Carlo.
- Understand the advantages and disadvantages of sampling-based methods.
- Know the standard deviation (i.e., the RMSE) of a Monte Carlo approximation, and know that it represents the rate of convergence.
- Be able to derive the basic properties of simple Monte Carlo approximations: consistency, unbiasedness, variance, standard deviation, RMSE.
- Be able to identify the limit of a given Monte Carlo approximation.
- Know how to construct a Monte Carlo approximation for: posterior probabilities, posterior densities, posterior expected loss, posterior predictive distribution, marginal likelihood.

- Know the condition under which a simple Monte Carlo approximation is consistent.
- Be able to give an example for which a simple Monte Carlo approximation is not consistent.
- Know that the harmonic mean approximation is consistent, but performs poorly.

9 Importance sampling

(References: Chapter 5 (Monte Carlo approx), mathematicalmonk ML 17.5–17.7)

- Be able to derive the basic importance sampling approximation.
- Know how to construct a basic importance sampling approximation for a given expectation and a given proposal distribution.
- Be able to identify the limit of a given importance sampling approximation.
- Understand the advantages and disadvantages of importance sampling compared to simple Monte Carlo.
- Be able to derive the basic properties: consistency, unbiasedness, variance, standard deviation, RMSE.
- Understand the properties that the proposal distribution must have in order to obtain a consistent importance sampling approximation.
- Understand how to choose the proposal distribution, in order to minimize the approximation error.
- Know that it is possible to handle unknown normalization constants.

10 Basic techniques for generating samples

(References: Chapter 5 (Monte Carlo approx), mathematicalmonk ML 17.8–17.14)

- Be able to apply the inverse c.d.f. method (derive a formula for transforming Uniform$(0,1)$ samples into samples from the desired distribution), given the p.d.f. of the desired distribution, possibly with an unknown normalization constant.
- Know that the c.d.f. of a continuous distribution transforms samples from that distribution into Uniform$(0,1)$ samples.
- Know the precise statement of the inverse c.d.f. method, and understand why it is necessary to use the generalized inverse.
- Understand the rejection principle, and be able to show that it is true.
- Understand the projection principle, and be able to show that it is true.
- Understand the rejection sampling procedure, and be able to show that it is true.

11 Gibbs sampling

(References: Chapter 6 (Gibbs sampling), Hoff chapter 6)

- Know the basic Gibbs sampling algorithm for a distribution $p(\theta_1, \ldots, \theta_k)$.
- Be able to derive a Gibbs sampler for a given distribution.
- Know the following terminology: full conditional distribution, sweep/scan, Gibbs updates, burn-in period, good mixing.
- Understand the advantages and disadvantages of MCMC.
- Know the definition of a Markov chain.
- Know how to use the output of an MCMC algorithm to approximating the following quantities:
  - posterior expectations
  - posterior probability of a given event
  - posterior densities and c.d.f.s
  - posterior predictive density and c.d.f.
- Understand why a burn-in period is often needed.
- Understand the utility of semi-conjugate (a.k.a. conditionally-conjugate) priors in the context of Gibbs sampling.
- Know how to choose semi-conjugate priors for commonly-used models (e.g., mean and variance for normal distribution).
- Understand why Gibbs sampling is so useful when using hyperpriors and/or hierarchical models.
- Understand the concept of data augmentation / auxiliary variables.

12 Priors

(References: Chapters 3 & 6, Hoff chapter 9)

- Understand the following terms:
  - conjugate prior
  - semi-conjugate prior
  - weakly-informative prior
  - unit-information prior
  - g-prior
  - data-dependent prior
  - improper prior
  - non-informative prior

- Understand how the "posterior" is defined when using an improper prior (see Chapter 6).
- Be able to choose prior parameter values that appropriately represent your prior beliefs.

13 MCMC diagnostics

(References: Chapter 6 (Gibbs sampling), Hoff section 6.6)

- Know how to construct and interpret the following MCMC diagnostics:
  - traceplots
  - running averages
  - autocorrelation function (ACF)
  - scatterplots
  - estimated posterior densities
- Understand the limitations of MCMC diagnostics (we can only tell when things are going wrong, not when things are going right).
- Be able to give a specific example in which mixing will appear to be fine according to standard diagnostics, but mixing will actually be very poor.
- Understand why a change of variables can sometimes significantly improve MCMC mixing.
- Understand why MCMC algorithms sometimes mix poorly on distributions with multiple modes.

14 Multivariate normal/Gaussian distribution

(References: Hoff chapter 7, mathematicalmonk PP 6.1–6.10)

- Know the density (p.d.f.) of the multivariate normal distribution.
- Know the definitions of:
  - the covariance and (Pearson's) correlation coefficient between two univariate random variables.
  - the covariance matrix and the precision matrix.
  - symmetric positive definite matrix (know at least two definitions).
  - matrix determinant (know at least two definitions).
  - matrix inverse.
- Know how to parameterize the covariance matrix of a bivariate normal in terms of the standard deviations σ_1, σ_2 and the correlation coefficient ρ.
- Know the relationship between independence and zero correlation among entries of a multivariate normal vector.

- Know the affine transformation property.
- Know how to construct a multivariate normal random vector from i.i.d. univariate standard normals.
- Know how to transform a multivariate random normal vector into i.i.d. univariate standard normals ("sphering").
- Know how to put a semi-conjugate prior on the mean and covariance matrix.
- Know how to put a semi-conjugate prior on the mean and precision matrix.
- Be able to derive the full conditionals for the mean and covariance (or precision) when using a semi-conjugate prior.
- Recognize the multivariate normal density as exp(quadratic).

15 Conditional independence relationships and graphical models

(References: Bishop chapter, mathematicalmonk ML 13.1–13.9)

- Understand that a graphical model does not actually specify a model, but rather a set of conditional independence properties.
- Understand why graphical models are useful.
- Know how to write down a directed graphical model (DGM) for a given probabilistic model.
- Know how to write down the factorization of the joint distribution specified by a DGM.
- Know what it means for a distribution to respect a given DGM.
- Understand that one cannot determine *dependence* from a graphical model, only independence.
- Be able to write down a DGM that is respected by any distribution on 5 variables.
- Be able to determine the moral graph associated with a given DGM.
- Know what it means for a distribution to respect a given undirected graphical model (UGM).
- Know how to determine conditional independence relationships from a UGM.

16 Group comparisons and hierarchical models

(References: Hoff chapter 8)

- Understand how to define a hierarchical model for data from several groups.
- Be able to write down an appropriate hierarchical model, given only a verbal description of the data.

- Be able to give examples in which a hierarchical model would be useful.
- Understand the concept of "sharing statistical strength" (when using a hierarchical model, data from one group helps to infer the parameters of another group).
- Be able to derive a Gibbs sampler for a hierarchical model with semi-conjugate priors (including, but not limited to, the hierarchical normal models described in Hoff 8.3 and 8.5).
- Understand how deFinetti's theorem can be used to justify modeling the observations with each group, as well as the group parameters, as conditionally i.i.d.
- Understand the concept of shrinkage, and why it can improve performance.

17 Linear regression

(References: Hoff chapter 9, mathematicalmonk ML 9.1–10.7)

- Be able to write down the normal linear regression model.
- Be able to identify when it makes sense to use linear regression in a given problem.
- Understand how basis functions can be used to model non-linear relationships between the predictor variables x_i and outcomes y.
- Understand what is "linear" about linear regression (the mean $\mathbb{E}(Y|x)$ is a linear function of the parameters β, but not necessarily of the predictor variables x_i).
- Be able to derive the ordinary least squares (OLS) estimator of the coefficients (i.e., the maximum likelihood estimator).
- Know how to define semi-conjugate priors for the coefficients β and the variance σ^2 (or precision λ).
- Know how to define a g-prior for the coefficients.
- Know how to define a unit-information prior for the coefficients.
- Be able to derive the resulting full conditionals, for each of these types of priors (note that the g-prior and unit-information are essentially special cases of the semi-conjugate prior).

18 Bayesian hypothesis testing and model selection/inference

(References: Lecture notes on course website, Hoff section 9.3, mathematicalmonk ML 12.1–12.4)

- Understand the concept of Bayesian hypothesis testing, and how it differs from frequentist hypothesis testing.
- Know that Bayesian hypothesis testing, Bayesian model selection, and Bayesian model averaging are all the same thing, mathematically.

- Be able to give examples of problems in which Bayesian hypothesis testing would make sense.
- Given a verbal description of a problem, be able to identify when Bayesian hypothesis testing would make sense, and write down an appropriate model.
- In simple cases, be able to compute the posterior on hypotheses analytically.
- Understand how decision theory can be used, if one hypothesis must be selected.
- Know the definition of Bayes factors, and how to interpret them.
- Know that "posterior odds = Bayes factor times prior odds".
- Be able to compute Bayes factors in simple cases.
- Understand how Bayes factors (and the posterior on hypotheses) can be strongly affected by the prior on parameters. Understand Lindley's "paradox".
- Understand why improper priors cannot be used when doing Bayesian hypothesis testing. (An improper prior is only defined up to a multiplicative constant.)
- Know that Bayes factors can be non-monotone in the sample size, and understand why this is the case.

19 Bayesian variable selection

(References: Hoff section 9.3)

- Understand why it can be advantageous to use a subset of variables, instead of all of them.
- Understand the basic idea of the backwards elimination procedure (but not necessarily the details).
- Understand how variable selection can be viewed as a hypothesis testing problem.
- Be able to write down a model for Bayesian variable selection.
- Be able to derive a sampler for the posterior, including a Gibbs sampler for the indicator variables (indicating which coefficients are included).

20 Metropolis–Hastings MCMC

(References: Hoff chapter 10, mathematicalmonk ML 18.1–18.9)

- Be able to write down the Metropolis algorithm and the Metropolis–Hastings (MH) algorithm.
- Be able to explain some of the advantages/disadvantages of MH versus Gibbs sampling.
- Be able to show that both the Metropolis algorithm and Gibbs sampling are special cases of MH.
- Understand the intuition behind the acceptance ratio in the Metropolis algorithm.

- Understand the rough idea behind the acceptance ratio in the MH algorithm.
- Understand why the choice of proposal distribution affects how well the resulting Markov chain mixes.
- In simple cases, be able to choose a reasonable proposal distribution.
- In simple cases, be able to derive the density of the proposal distribution (in order to compute the acceptance ratio).

21 Markov chains and combining MCMC moves

(References: Hoff chapter 10, mathematicalmonk ML 14.1–14.3, 18.1–18.9)

- Know the mathematical definition of a Markov chain (MC). "The Markov property"
- Know the mathematical definition of the following terms, for a time-homogeneous MC taking values in a discrete (countable) space:
  1. transition matrix
  2. irreducible MC
  3. aperiodic MC
  4. stationary distribution (a.k.a. invariant distribution)
  5. detailed balance
- Be able to give examples of MCs that have each of the following properties, and examples that do not: is irreducible, is aperiodic, has a stationary distribution.
- In simple cases, be able to identify whether an MC is irreducible and/or aperiodic, and be able to identify a stationary distribution (if one exists).
- Know what the ergodic theorem is, and how it is used in MCMC.
- Be able to show that having detailed balance implies that the target distribution is a stationary distribution.
- Understand the intuition behind the definition of stationary distribution, and detailed balance (e.g., the water flow analogy).
- Be able to show that the MH algorithm induces an MC having detailed balance (for the target distribution).
- Know how moves (transition matrices) can be combined, using
  1. a product of transition matrices, i.e., a repeating cycle of moves (e.g., as in "fixed-scan" Gibbs sampling),
  2. a mixture of transition matrices, i.e., a random choice of move (e.g., as in "random-scan" Gibbs).
- Understand why the choice of move (generally-speaking) cannot depend on the state of the MC.

- Understand how MH can be used within Gibbs sampling (MH-within-Gibbs).
- Be able to derive an MH-within-Gibbs sampler for a given model.

22 Advanced Monte Carlo methods

- In *very* general terms, know the basic idea of the following:

  1. quasi-Monte carlo
  2. slice sampling
  3. Hamiltonian MC, a.k.a. hybrid MC
  4. approximate Bayesian computation (ABC)

23 Mixture models

(References: mathematicalmonk ML 16.6)

- Know the definition of a mixture.
- Be able to write down a mixture model for a given problem.
- Know that mixture models can be used for density estimation, clustering, and latent structure modeling.
- Know how to derive a Gibbs sampler for a two-component finite mixture model (this was covered in Chapter 6).