ECE440 - Midterm Review Notes

Probability review

You should be fluent with notions such as probability spaces, the axioms of probability and their consequences, the law of total probability and Bayes' rule, discrete and continuous random variables (RVs), expectations, independence, conditional probability and expectations, and notions of convergence of sequences of RVs.

Some noteworthy concepts and results you should make sure you are comfortable with are:

1) E and F are *mutually exclusive* (disjoint) events if $E \cap F = \emptyset$. From the axioms of probability this implies that for disjoint E and F, then $\mathrm{P}[E \cup F] = \mathrm{P}[E] + \mathrm{P}[F]$.
2) E and F are *independent* events if $\mathrm{P}[E \cap F] = \mathrm{P}[E]\,\mathrm{P}[F]$.
3) You should be able to tell the difference between discrete and continuous RVs.
4) A *Bernoulli* RV X with parameter p indicates that a random event E occurred (a "success"), where $\mathrm{P}[E] = p$. Recall the pmf, that $\mathbb{E}[X] = p$ and $\mathrm{var}[X] = p(1-p)$.
5) A *Geometric* RV X with parameter p counts the number of independent Bernoulli trials till we register the first success. Recall the pmf, that $\mathbb{E}[X] = 1/p$ and $\mathrm{var}[X] = (1-p)/p^2$.
6) A *Binomial* RV X with parameters n and p counts the number of successes in n independent Bernoulli trials. Recall that $\mathbb{E}[X] = np$ and $\mathrm{var}[X] = np(1-p)$. For $B_1, \ldots, B_n$ i.i.d. Bernoulli RVs with parameter p, one can write
$$X = \sum_{i=1}^{n} B_i.$$
7) A *Poisson* RV X with parameter λ counts rare events or "arrivals". Recall the pmf, that $\mathbb{E}[X] = \lambda$ and $\mathrm{var}[X] = \lambda$. If $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ are independent, then $Y = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$. The law of rare events asserts that the distribution of $X \sim \text{Binomial}(n, p)$ converges to a $\text{Poisson}(\lambda)$ as $n \to \infty$, provided $np = \lambda$.
8) A *Gaussian* (Normal) RV X with parameters μ and σ^2 models randomness arising from the superposition of large number of random effects. (This statement should be clear if you understood the Central Limit Theorem.) Recall the pdf, that $\mathbb{E}[X] = \mu$ and $\mathrm{var}[X] = \sigma^2$.
9) An *indicator* RV $\mathbb{I}\{E\}$ indicates the occurrence of event E. Indicator RVs are Bernoulli distributed, with parameter $p = \mathrm{P}[E]$. You should be fluent in operating with indicator RVs. The expectation of an indicator RV is the probability of the indicated event, i.e., $\mathbb{E}[\mathbb{I}\{E\}] = \mathrm{P}[E]$.
10) It is useful to master the properties of expected values, such as the "law of the unconscious statistician" to calculate the expected value of a function of a RV, i.e., $\mathbb{E}[f(X)]$. The expected value is a linear operator, hence for scalars a_i and b_i, $i = 1, \ldots, n$, then
$$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i] + b_i.$$
In general, expectation can be interchanged with any linear operation. If X and Y are independent RVs, then for any functions $f(\cdot)$ and $g(\cdot)$ one has
$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)].$$

Expectation can be interchanged with products only if the RVs are independent, but it can always be interchanged with sums. Also, for an independent sequence of RVs $X_1, \ldots, X_n$ and scalars a_i and b_i, $i = 1, \ldots, n$, then

$$\text{var}\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 \text{var}\left[X_i\right].$$

11) The covariance is a measure of linear dependence between pairs of RVs, given by

$$\text{cov}[X, Y] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right] = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right].$$

Recall that independent RVs are also uncorrelated, but the converse is not true in general (it is true though, if the random variables are jointly Gaussian).

12) Make sure you understand the statements of Markov's and Chebyshev's inequalities.
13) You should have an intuitive understanding of the difference between almost sure convergence, convergence in probability and convergence in distribution. Recall that the limit of a sequence of RVs is itself a RV.
14) The simplest class of random processes are *i.i.d. sequences of RVs*. You should understand and be able to apply the two main limiting theorems that pertain to sums of i.i.d. RVs. The *Strong Law of Large Numbers* asserts that for a sequence $X_{\mathbb{N}} = X_1, X_2, \ldots$ of i.i.d. RVs with mean $\mathbb{E}\left[X_1\right] = \mu$, the sample average converges to the mean almost surely, i.e.,

$$\mathrm{P}\left[\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu\right] = 1.$$

If the i.i.d. RVs have variance $\mathbb{E}\left[(X_1 - \mu)^2\right] = \sigma^2$, the *Central Limit Theorem asserts* that

$$\lim_{n\to\infty} \mathrm{P}\left[\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \le x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du.$$

This implies the above normalized sum of RVs converges in distribution to a standard Gaussian RV.

15) Conditional probabilities and expectations and very useful tools for the analysis of random processes. Make sure you are comfortable with the definitions and examples from the lecture slides, and that you understand the related problems from the homework assignments.
16) The conditional expectation $E[X \mid Y]$ is a function of Y, hence a RV which takes the values $E[X \mid Y = y]$. The law of iterated expectations states that $\mathbb{E}\left[X\right] = \mathbb{E}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right]$, and is often helpful in computing expected values by conditioning. Likewise, one can compute the variance by using the formula

$$\text{var}\left[X\right] = \mathbb{E}_Y\left[\text{var}_X\left[X \mid Y\right]\right] + \text{var}_Y\left[\mathbb{E}_X\left[X \mid Y\right]\right].$$

Markov chains

You should be fluent with notions such as the Markov property, (multi-step) transition probabilities, representing a Markov chain (MC) through the transition probability matrix $\mathbf{P}$ or state transition diagram, the Chapman-Kolmogorov equations, computing n-step transition probabilities and unconditional probabilities, classifying states of a MC, identifying communicating classes and irreducible MCs, computing stationary distributions, and understanding as well as applying the main results on limiting distributions and long-run time averages of ergodic MCs.

Some noteworthy concepts and results you should make sure you are comfortable with are:

1) To claim that a random process $X_{\mathbb{N}} = X_1, X_2, \ldots$ is a MC, the direct approach is to show the Markov property holds, i.e., that

$$\mathrm{P}\left[X_{n+1} = j \,\middle|\, X_n = i, \mathbf{X}_{n-1} = \mathbf{x}\right] = \mathrm{P}\left[X_{n+1} = j \,\middle|\, X_n = i\right] = P_{ij}$$

only depends on i, j. Oftentimes, it may be easier to show that

$$X_{n+1} = f(X_n, D_{n+1})$$

where f is a function of the previous state and some i.i.d. sequence $D_{\mathbb{N}} = D_1, D_2, \ldots$, which is also independent of X_0.

2) The state space of a MC can be finite or infinite, but it should be countable.
3) Transition probabilities P_{ij} are non-negative and sum up to one, i.e., $P_{ij} \geq 0$ for all i, j, and $\sum_{j=1}^{\infty} P_{ij} = 1$ for all i. Equivalently, all *row-wise* sums of matrix $\mathbf{P}$ equal 1, and all per-state *outgoing* arrow weights in a state transition diagram sum up to 1.
4) The Chapman-Kolmogorov equations are

$$P_{ij}^{n+m} = \sum_{k=1}^{\infty} P_{ik}^{m} P_{kj}^{n}.$$

You should understand the arguments used to derive them, and their representation in matrix form, namely $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)}\mathbf{P}^{(n)}$. A direct consequence of great practical value is that the matrix of n-step transition probabilities is given by the n-th power of $\mathbf{P}$, i.e., $\mathbf{P}^{(n)} = \mathbf{P}^{n}$.

5) You should be fluent in calculating (multi-step) transition probabilities, unconditional probabilities once the initial distribution or the unconditional distribution of an earlier time step is specified, joint probabilities, and expectations involving states of a MC.
6) States of a MC can be transient or recurrent. If state i is transient, there is a strictly positive probability that given the MC is currently in state i, it will never return to i in the future. If state j is recurrent, the MC revisits j infinitely often. You should be familiar with the different characterizations of transience and recurrence used to classify states of a MC.
7) You should train yourself to quickly identify communicating classes in a MC. If there is a single class, the MC is termed irreducible. Because transience and recurrence are class properties, all states in a given communicating class are either transient or recurrent.
8) In a finite MC, at least one state must be recurrent. (Make sure you understand why this is the case.) A finite irreducible MC must be recurrent (all states are recurrent). An infinite irreducible MC may be transient (all states are transient); e.g., the biased random walk.
9) An absorbing state always defines a communicating class, of which it is the only member. Absorbing states are always recurrent.
10) You should be comfortable with the notion of (a)periodic state. Periodicity is a class property. If $P_{ii} \neq 0$ for some state i, then all states in the class i belongs to are aperiodic. (First, check the diagonal of $\mathbf{P}$.) In general, the period d is given by $d = \gcd\{n : P_{ii}^{n} \neq 0\}$.
11) State i is recurrent if an only if the return time $T_i := \min\{n > 0 : X_n = i\}$ is finite almost surely, i.e.,

$$\mathrm{P}\left[T_i = \infty \,\middle|\, X_0 = i\right] = 0.$$

Recurrent states can be either positive recurrent ($\mathbb{E}\left[T_i \,\middle|\, X_0 = i\right] < \infty$), or null recurrent ($\mathbb{E}\left[T_i \,\middle|\, X_0 = i\right] = \infty$). Positive and null recurrence are class properties. All recurrent states in a finite MC must be positive recurrent. See Homework 6, Problem 6 for an example of a null-recurrent MC.

12) The best behaved MCs are ergodic, which are irreducible, positive recurrent and aperiodic. For an ergodic MC, the limiting probabilities $\lim_{n\to\infty} P_{ij}^n$ exist (converge) and are independent of the initial state, i.e.,

$$\pi_j = \lim_{n\to\infty} P_{ij}^n.$$

The steady-state probabilities $\pi_j \geq 0$ are the unique nonnegative solution of the system of linear equations

$$\pi_j = \sum_{n=1}^{\infty} \pi_i P_{ij}, \qquad \sum_{j=1}^{\infty} \pi_j = 1.$$

13) Limiting distributions π_i are known as stationary distributions, because if the MC is initialized with $\mathrm{P}\left[X_0 = i\right] = \pi_i$, then $\mathrm{P}\left[X_n = i\right] = \pi_i$ for all $n > 0$, and all i.
14) The Ergodic Theorem states that the long-run fraction of time an ergodic MC spends in state i converges to the stationary (limiting) probability π_i. Specifically, if $X_{\mathbb{N}} = X_1, X_2, \ldots$ is an ergodic MC with stationary probabilities π_i, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\left\{X_m = i\right\} = \pi_i, \quad \text{a.s.}$$

You should understand the conceptual difference between ensemble average and ergodic (time) average. The first is an average over all possible realizations of the random process, the second one is a time average over a single realization of the process. Ergodic MCs are thus nice because a single realization provides as much information as if we were able to observe all realizations.
15) Limiting probabilities of transient or null-recurrent states converge to 0. Limiting probabilities of periodic states do not converge, they oscillate. But still the long-run fraction of time spent in a periodic state converges, i.e. the Ergodic Theorem still applies.
16) Make sure you understand the asymptotic ($n \to \infty$) behavior of a general reducible MC.

A few common mistakes made in the homework problems

We encountered some systematic mistakes made in the homework problems. Make sure you look at the solutions to understand where the error was, and the right way to solve it.

- **Homework 3, Problem 4:** In this problem, some of you incorrectly wrote

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[\sum_{i=1}^{N} T_i\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[T_i\right].$$

The last equality does not hold in this case, because the RVs N and T_i are not independent. In particular notice that if $T_i = 2$, then this implies that $N = i$.
- **Homework 3, Problem 5:** Given $\mathbb{E}\left[X\right] = \mu$ and $\mathrm{var}\left[X\right] = \mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$, some of you incorrectly wrote $\mathbb{E}\left[X^2\right] = \sigma^2$ or $\mathbb{E}\left[X^2\right] = \mu^2$, instead of $\mathbb{E}\left[X^2\right] = \sigma^2 + \mu^2$.

- **Homework 3, Problem 6, Part B:** Make sure you understand the answer to this part, in particular why convergence in distribution does not imply that realizations of the RVs should be approximately equal.
- **Homework 4, Problem 6:** Some of you struggled to find the requested example. Please check the solution.
- **Homework 5, Problem 3:** This is a challenging yet very important problem. Make sure you understand the arguments used in the solution.

1. Suppose A, B, C are events with $\mathrm{P}[C] > 0$ and $\mathrm{P}[B \cap C] > 0$. Derive a simple expression for $\mathrm{P}\left[A \cap B \,\middle|\, C\right]$ in terms of $\mathrm{P}\left[A \,\middle|\, B \cap C\right]$ and $\mathrm{P}\left[B \,\middle|\, C\right]$. Show your work.

2. Suppose X and Y are random variables with joint probability mass function given by

$$\begin{aligned}
\mathrm{P}[X = 1, Y = 1] &= 1/12, \\
\mathrm{P}[X = 1, Y = 2] &= 3/12, \\
\mathrm{P}[X = 2, Y = 1] &= 2/12, \\
\mathrm{P}[X = 2, Y = 2] &= 6/12.
\end{aligned}$$

(a) $\mathrm{P}\left[X + Y = 2 \,\middle|\, X = 1\right] = ?$

(b) Are X and Y independent? Justify your answer.

(c) Are X and Y identically distributed? Justify your answer.

(d) $\operatorname{var}[3X - 1] = ?$

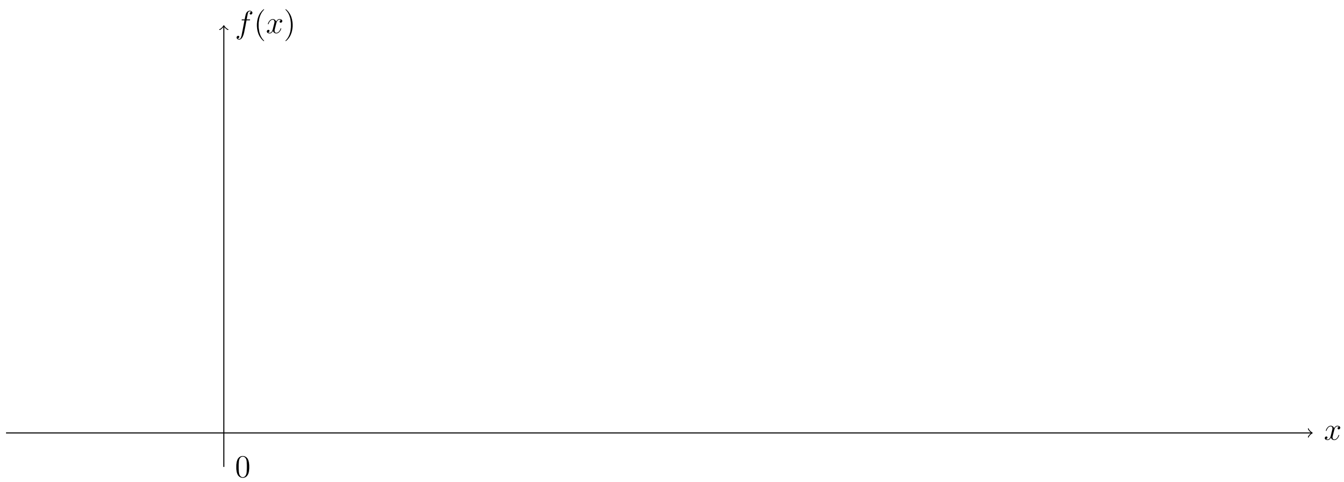
3. A professor continually gives exams to her students. She can give three possible types of exams, and her class is graded as either having done well or badly. Let p_i denote the probability that the class does well on a type i exam, and suppose that $p_1 = 0.3$, $p_2 = 0.6$ and $p_3 = 0.9$. If the class does well on an exam, then the next exam is equally likely to be any of the three types. If the class does badly, then the next exam is always type 1. In the long run, what proportion of exams are type 2?

4. Suppose that $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$ is an i.i.d. sequence of random variables with mean $\mathbb{E}[X_1] = 5$ and variance $\text{var}[X_1] = 1$. Consider the following random variables

$$S_n := \sum_{i=1}^{n} X_i,$$
$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n},$$
$$Z_n := \frac{\sum_{i=1}^{n} X_i - 5n}{\sqrt{n}} = \frac{S_n - 5n}{\sqrt{n}}.$$

Suppose that $n = 100$ is large enough so that limiting behaviors become apparent.

(a) Sketch the probability density functions (pdfs) of S_{100}, $\bar{X}_{100}$, and Z_{100}, superimposing the three plots in the set of axes provided below. Only rough, qualitative depictions are required, focusing on the notable values where the pdfs are centered, and their relative widths and heights. Justify your answer and label your plots.



(b) Will your plots fundamentally change if the common distribution of the random variables $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$ differs from that of (a), while the mean and variance remain the same (that is, one still has $\mathbb{E}[X_1] = 5$ and $\text{var}[X_1] = 1$)? Explain.

5. Old McDonald had a farm, but now he runs a monster-sighting business in Loch Ness, Scotland. Every day, he is unable to run the boat tour due to bad weather with probability p, independently of all other days. McDonald works every day except the bad-weather days, which he takes as holiday.

Let Y be the number of consecutive days McDonald has to work between bad-weather days. Let X be the total number of customers who go on McDondald's boat trip in this period of Y days. Conditioned on Y, the distribution of X is $X \,\big|\, Y \sim \text{Poisson}(\lambda Y)$, meaning the conditional probability mass function is given by

$$\text{P}\left(X = x \,\middle|\, Y = y\right) = \frac{(\lambda y)^x e^{-\lambda y}}{x!}.$$

(a) $\mathbb{E}\left[Y\right] = ?$ [Hint: Argue that the random variable $Z := Y + 1 \sim \text{Geometric}(p)$]

(b) $\mathbb{E}\left[X \,\middle|\, Y = y\right] = ?$

(c) $\mathbb{E}[X] = ?$