MATH 341: PROBABILITY: FALL 2009
REVIEW SHEET

STEVEN J. MILLER (SJM1@WILLIAMS.EDU)

ABSTRACT. Below is a summary of definitions and some key lemmas and theorems from Math 341.

---

1. CHAPTER 1

1.1. Definitions.

We record some common definitions below.

- **Sample Space (Ω):** all possible outcomes. Example: toss coin thrice: $\{HHH, \dots, TTT\}$; toss until get head: $\{H, TH, TTH, \dots\}$.

- **Events:** Subsets of sample space Ω. Example: at least 2 of 3 tosses a head: $\{HHT, HTH, THH, HHH\}$.

- **Complement:** $A^c = \Omega - A$.

- **Field:**
  $\diamond$ $A, B \in \mathcal{F}$ then $A \cup B$ and $A \cap B$ in $\mathcal{F}$.
  $\diamond$ $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
  $\diamond$ $\varphi \in \mathcal{F}$ (so $\Omega \in \mathcal{F}$).
  $\diamond$ if also $A_i \in \mathcal{F}$ implies $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ then a σ**-field**.

- **Finitely additive:** disjoint union then $\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$; **countably additive** if the $\{A_i\}$ pairwise disjoint implies $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

- **Probability space:** A triple $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space if Ω is a sample space with σ-field $\mathcal{F}$ and a **probability measure** $\mathbb{P}$ satisfying
  $\diamond$ $\mathbb{P}(\varphi) = 0, \mathbb{P}(\Omega) = 1$.
  $\diamond$ $\mathbb{P}$ is countably additive: for a disjoint union, $\mathbb{P}(\cup_{i=1}^{\infty}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

- **Conditional probability:** If $\mathbb{P}(B) > 0$ then the conditional probability of A occurring given B, denoted $\mathbb{P}(A|B)$, is

$$\mathbb{P}(A|B) \;=\; \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Date: October 10, 2009.

◇ Interpretation through counting:

$$\frac{N(A\cap B)}{N(B)} \ = \ \frac{N(A\cap B)/N}{N(B)/N} \ \to \ \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$

◇ **Example:** roll fair die twice: what is probability of a 7 or an 11 given first roll is 3? Ans: $\frac{1/36}{6/36} = 1/6$ and $\frac{0/36}{6/36} = 0$.

- **Partition** A family of events $B_1, \dots, B_n$ is a partition of Ω if the $\{B_i\}$'s are disjoint and $\cup_{i=1}^n B_i = \Omega$.

- **Independence** A and B are **independent** if
$$\mathbb{P}(A\cap B) \ = \ \mathbb{P}(A)\mathbb{P}(B).$$
More generally, a family $\{A_i\}_{i\in I}$ is independent if
$$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) \ = \ \prod_{i\in J}\mathbb{P}(A_i) \quad \text{for any } J\subset I.$$

1.2. **Basic Lemmas.**

Lemma 1.1. *For a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *we have*

- *Law of total probability:* $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

- $A\subset B$ *implies* $\mathbb{P}(A) \le \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A)$.

- $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cap B)$.

- $\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j}\mathbb{P}(A_i\cap A_j) + \cdots + (-1)^{n+1}\mathbb{P}(A_1\cap\cdots\cap A_n)$ *(Inclusion - Exclusion Principle).*

Lemma 1.2. $A_1\subset A_2\subset\cdots$ *and* $B_1\supset B_2\supset\cdots$, *then*

- *If* $A = \cup_{i=1}^\infty A_i$ *then* $\mathbb{P}(A) = \lim_{n\to\infty}\mathbb{P}(\cup_{i=1}^n A_i)$.

- *If* $B = \cap_{i=1}^\infty B_i$ *then* $\mathbb{P}(B) \ = \ \lim_{n\to\infty}\mathbb{P}(\cap_{i=1}^n B_i)$.

Lemma 1.3. *If* $0 < \mathbb{P}(B) < 1$ *then for any event* A *we have*
$$\mathbb{P}(A) \ = \ \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c).$$
If the $\{B_i\}$ *form a pairwise disjoint partition, then*
$$\mathbb{P}(A) \ = \ \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

2. Chapter 2

2.1. Definitions.

- **Random Variables:** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A random variable is a function X from the sample space Ω to the real numbers with the property that $\{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F}$ for each x.

  $\diamond$ **Example:** Ω: tosses of a fair coin five times, $\mathcal{F} = 2^{\Omega}$, the set of all subsets of Ω, and let $X(\omega)$ denote the number of heads in ω. As there are $2^5 = 32$ elements, there are 2^{32} or about 4,000,000,000 elements in $\mathcal{F}$. Each element of $\mathcal{F}$ is a subset of Ω, and each subset of Ω is an element of $\mathcal{F}$. If we write $F = \{\omega_1, \dots, \omega_k\}$ for an element of $\mathcal{F}$, then $\mathbb{P}(F) = \sum_{i=1}^{k} \mathbb{P}(\omega_i)$. A straightforward computation shows that X has the desired property; this is clear as all subsets of Ω are in $\mathcal{F}$! If $x = 1$ then $\{\omega \in \Omega : X(\omega) \le 1\} = \{TTTTT, TTTTH, TTTHT, TTHTT, THTTT, HTTTT\}$. If instead we took $x = 4$, then the set would be all outcomes except $HHHHH$.

- **Distribution Function:** The distribution function of a random variable $X : \Omega \to \mathbb{R}$ is the function $F : \mathbb{R} \to [0,1]$ given by $F(x) = \mathbb{P}(X \le x)$. In other words, it's the probability of observing a value of X of at most x.

  $\diamond$ **Example:** Consider five tosses of a fair coin. We have $F(0) = 1/32$, $F(1) = 6/32$, $F(2) = 16/32$, $F(3) = 26/32$, $F(4) = 31/32$ and $F(5) = 32/32$. Our function is supposed to be defined for all real x, so what we really have is the following: $F(x) = 0$ if $x < 0$, $F(x) = 1/32$ if $0 \le x < 1$, $F(x) = 6/32$ if $1 \le x < 2$, and so on.

- **Discrete Random Variables:** A random variable X is discrete if it takes values in a countable subset $\{x_1, x_2, \dots\}$ of $\mathbb{R}$. It has probability mass function $f : \mathbb{R} \to [0,1]$ given by $f(x) = \mathbb{P}(X = x)$.

  $\diamond$ **Example:** Toss a fair coin until the first head is obtained. Then $\Omega = \{H, TH, TTH, \dots\}$. Let X be the number of tosses needed to obtain the first head. Then X is discrete, taking on the values $\{1, 2, 3, \dots\}$, with the probability X equals n just $1/2^n$.

- **Continuous Random Variables:** A random variable X is continuous if its distribution function can be written as $F(x) = \int_{-\infty}^{x} f(u)du$ for some integrable function f (which is called the probability density function of X).

  $\diamond$ **Example:** Let $\Omega = [0,1]$ and let $\mathcal{F}$ be the σ-field generated by the open intervals. (This is the standard σ-field.) Let $X(\omega)$ equal ω^2. If we let Y be uniformly distributed on $[0,1]$, then we see $\mathbb{P}(X \le x)$ is the same as $\mathbb{P}(Y \le \sqrt{x})$, which is just $\sqrt{x}$. We are therefore looking for f so that $\sqrt{x} = \int_0^x f(u)du$ for $0 \le x \le 1$. Differentiating both sides gives $\frac{1}{2}x^{-1/2} = f(x)$ (note the integral is $\mathfrak{F}(x) - \mathfrak{F}(0)$ with $\mathfrak{F}$ any anti-derivative of f; differentiating yields the claim as $\mathfrak{F}' = f$). We see that for our random variable X, we may take $f(u) = 1/2\sqrt{u}$ for $0 < u \le 1$ and 0 otherwise.

- **Joint Distribution of a Random Vector:** The joint distribution function of a random vector $\overrightarrow{\mathbf{X}} = (X_1, \dots, X_n)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the function $F_{\overrightarrow{\mathbf{X}}} : \mathbb{R}^n \to [0,1]$ given by $F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}) = \mathbb{P}(\overrightarrow{\mathbf{X}} \leq \overrightarrow{x})$ for $\overrightarrow{x} \in \mathbb{R}^n$, where $\overrightarrow{x} \leq \overrightarrow{y}$ means each $x_i \leq y_i$, and $\{\overrightarrow{\mathbf{X}} \leq \overrightarrow{x}\} = \{\omega \in \Omega : \overrightarrow{\mathbf{X}}(\omega) \leq \overrightarrow{x}\}$.

- **Jointly Discrete** $X_1, \dots, X_n$ random vectors on $(\Omega, \mathcal{F}, \mathbb{P})$ are jointly discrete if $\overrightarrow{\mathbf{X}} = (X_1, \dots, X_n)$ takes values in a countable subset of $\mathbb{R}^n$ and has joint probability mass function $f : \mathbb{R}^n \to [0,1]$ given by
$$f(x_1, \dots, x_n) \ = \ \mathbb{P}(X_1 = x_1, \dots, X_n - x_n).$$

- **Jointly Continuous** Jointly continuous defined analogously, with
$$F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}) \ = \ \int_{u_1=-\infty}^{x_1} \cdots \int_{u_n=-\infty}^{x_n} f(u_1, \dots, u_n) du_1 \cdots du_n$$
for some integrable function $f : \mathbb{R}^n \to [0, \infty)$.

- **Marginals:** Same set-up as above, the j^{th} marginal F_{X_j} is defined by
$$F_{X_j}(x_j) \ := \lim_{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n \to \infty} F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}).$$

2.2. **Lemmas.**

Lemma 2.1. *The (cumulative) distribution function satisfies the properties:*

- $\lim_{x_1, \dots, x_n \to -\infty} F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}) = 0$, $\lim_{x_1, \dots, x_n \to \infty} F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}) = 1$.
- *If* $\overrightarrow{x} \leq \overrightarrow{x}'$ *then* $F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}) \leq F_{\overrightarrow{\mathbf{X}}}(\overrightarrow{x}')$.
- $F_{\overrightarrow{\mathbf{X}}}$ *continuous from above.*

---

3. Chapters 3 and 4

3.1. **Definition.**

- **Probability Mass Function:** The Probability Mass Function of a discrete random variable X is a function $f : \mathbb{R} \to [0,1]$ given by $f(x) = \mathbb{P}(X = x)$.

- **Probability Density Function:** The Probability Density Function of a continuous random variable X is the f such that $F(x) = \int_{-\infty}^{x} f(u) du$.

- **Independence of events:** Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

  $\diamond$ As $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$, if $\mathbb{P}(B) > 0$ this is equivalent to $\mathbb{P}(A|B) = \mathbb{P}(A)$, or that knowledge of one happening does not affect knowledge of the other happening.

- **Independence of random variables:** Two random variables X and Y are independent if for all x, y:
  - $\diamond$ **Discrete case:** events $\{X = x\}$ and $\{Y = y\}$ are independent.
  - $\diamond$ **Continuous case:** events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent.

- **Expectation (mean value, average):** X random variable with density / mass function f_X, then expected value is
  - $\diamond$ **Discrete case:** $\mathbb{E}[X] := \sum_x x f_X(x)$ if sum converges absolutely.
  - $\diamond$ **Continuous case:** $\mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x) dx$ if integral converges absolutely.

- **Moments:** Let X be a random variable. We define
  - $\diamond$ k^{th} moment: $m_k := \mathbb{E}[X^k]$ (if converges absolutely).

- Assume X has a finite mean, which we denote by μ (so $\mu = \mathbb{E}[X]$). We define
  - $\diamond$ k^{th} centered moment: $\sigma_k := \mathbb{E}[(X - \mu)^k]$ (if converges absolutely).

- **Variance:** Call σ_2 the variance, write it as σ^2. Note $\sigma^2 = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

3.2. Lemmas.

Lemma 3.1. *Standard properties of the probability mass function:*

- $F(x) = \sum_{x_i \leq x} f(x_i)$, *and* $f(x) = F(x) - \lim_{y \to x^-} F(y)$.
- $\{x : f(x) \neq 0\}$ *is at most countable.*
- $\sum_i f(x_i) = 1$ *where* $\{x_1, x_2, \dots\}$ *is where* f *is non-zero.*

Lemma 3.2. *Standard properties of the probability density function:*

- $\int_{-\infty}^{\infty} f(x) dx = 1$.
- $\mathbb{P}(X = x) = 0$ *for all* $x \in \mathbb{R}$.
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$.

Lemma 3.3. *Let* $g, h : \mathbb{R} \to \mathbb{R}$ *and assume* X *and* Y *are independent random variables. Then* $g(X)$ *and* $h(Y)$ *are independent.*

Lemma 3.4 (Key results)**.** *Let X and Y be two random variables, and let $a, b \in \mathbb{R}$.*

- *Linearity:* $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

- *Independence:* X, Y *independent then* $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. *If RHS holds say uncorrelated.*

- *Variance:* $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$ *if uncorrelated. In general:*

$$\begin{aligned} \mathrm{CoVar}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{CoVar}(X_i, X_j). \end{aligned}$$