CSCI 1550: PROBABILISTIC METHODS IN CS (NOTES)

ALEXANDER W. LEE

1. Events and Probability

Definition. *A probability space has three components:*

1. *a sample space Ω, which is the set of all possible outcomes of the random process modeled by the probability space;*
2. *a family of sets $\mathcal{F}$ representing the allowable events, where each set in $\mathcal{F}$ is a subset of the sample space Ω; and*
3. *a probability function $\Pr : \mathcal{F} \to \mathbb{R}$ satisfying the following definition.*

Definition. *A probability function is any function $\Pr : \mathcal{F} \to \mathbb{R}$ that satisfies the following conditions:*

1. *for any event E, $0 \leq \Pr(E) \leq 1$;*
2. *$\Pr(\Omega) = 1$; and*
3. *for any finite or countably infinite sequence of pairwise mutually disjoint events $E_1, E_2, E_3, \ldots$,*

$$\Pr\left(\bigcup_{i\geq 1} E_i\right) = \sum_{i\geq 1} \Pr(E_i).$$

Lemma. *For any two events E_1 and E_2,*

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2).$$

Lemma (Union Bound)**.** *For any finite or countably infinite sequence of events $E_1, E_2, \ldots$,*

$$\Pr\left(\bigcup_{i\geq 1} E_i\right) \leq \sum_{i\geq 1} \Pr(E_i).$$

Lemma. *Let $E_1, \ldots, E_n$ be any n events. Then*

$$\begin{aligned}\Pr\left(\bigcup_{i=1}^{n} E_i\right) = &\sum_{i=1}^{n} \Pr(E_i) - \sum_{i<j} \Pr(E_i \cap E_j) + \sum_{i<j<k} \Pr(E_i \cap E_j \cap E_k) \\ &- \cdots + (-1)^{l+1} \sum_{i_1<i_2<\cdots<i_l} \Pr\left(\bigcap_{r=1}^{l} E_{i_r}\right) + \cdots .\end{aligned}$$

Definition. *Two events E and F are independent if and only if*

$$\Pr(E \cap F) = \Pr(E) \cdot \Pr(F).$$

More generally, events $E_1, E_2, \ldots, E_k$ are mutually independent if and only if, for any subset $I \subseteq [1, k]$,

$$\Pr\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \Pr(E_i).$$

Definition. *The conditional probability that event E occurs given that event F occurs is*

$$\Pr(E \mid F) = \frac{\Pr(E \cap F)}{\Pr(F)}.$$

The conditional probability is well-defined only if $\Pr(F) > 0$.

Theorem (Law of Total Probability)**.** *Let $E_1, E_2, \ldots, E_n$ be mutually disjoint events in the sample space Ω, and let $\bigcup_{i=1}^{n} E_i = \Omega$. Then*

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B \mid E_i) \Pr(E_i).$$

Theorem (Bayes' Law)**.** *Assume that $E_1, E_2, \ldots, E_n$ are mutually disjoint events in the sample space Ω such that $\bigcup_{i=1}^{n} E_i = \Omega$. Then*

$$\Pr(E_j \mid B) = \frac{\Pr(E_j \cap B)}{\Pr(B)} = \frac{\Pr(B \mid E_j) \Pr(E_j)}{\sum_{i=1}^{n} \Pr(B \mid E_i) \Pr(E_i)}.$$

Theorem. *For any $x \in \mathbb{R}$,*

$$1 - x \leq e^{-x}.$$

Equivalently,

$$1 + x \leq e^{x}.$$

2. Discrete Random Variables and Expectation

Definition. *A random variable X on a sample space Ω is a real-valued (measurable) function on Ω; that is, $X : \Omega \to \mathbb{R}$. A discrete random variable is a random variable that takes on only a finite or countably infinite numbers of values.*

Definition. *Two random variable X and Y are independent if and only if*

$$\Pr((X = x) \cap (Y = y)) = \Pr(X = x) \cdot \Pr(Y = y)$$

for all values x and y. Similarly, random variables $X_1, X_2, \ldots, X_k$ are mutually independent if and only if, for any subset $I \subseteq [1, k]$ and any values $x_i, i \in I$,

$$\Pr\left(\bigcap_{i \in I} (X_i = x_i)\right) = \prod_{i \in I} \Pr(X_i = x_i).$$

Definition. *The expectation of a discrete random variable X, denoted by $\mathbf{E}[X]$, is given by*

$$\mathbf{E}[X] = \sum_{i} i \Pr(X = i),$$

where the summation is over all values in the range of X. The expectation is finite if $\sum_i |i| \Pr(X = i)$ converges; otherwise, the expectation is unbounded.

Theorem (Linearity of Expectations)**.** *For any finite collection of discrete random variables $X_1, X_2, \ldots, X_n$ with finite expectations,*

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

Lemma. *For any constant c and discrete random variable X,*

$$\mathbf{E}[cX] = c\mathbf{E}[X].$$

Definition. *Suppose that we run an experiment that succeeds with probability p and fails with probability $1-p$. A variable X is called a* Bernoulli *or an* indicator *random variable if*

$$X = \begin{cases} 1 & \text{if the experiments succeeds,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbf{E}[X] = \Pr(X = 1) = p$.

Definition. *Consider a sequence of n independent experiments, each of which succeeds with probability p. If we let X represent the number of successes in the n experiments, then X has a binomial distribution. A binomial random variable X with parameters n and p, denoted by $B(n,p)$, is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:*

$$\Pr(X = j) = \binom{n}{j} p^j (1-p)^{n-j}.$$

That is, the binomial random variable X equals j when there are exactly j successes and $n-j$ failures in n independent experiments, each of which is successful with probability p.

Lemma. *For a binomial random variable X with parameters n and p,*

$$\mathbf{E}[X] = np.$$

Definition.

$$\mathbf{E}[Y \mid Z = z] = \sum_{y} y \Pr(Y = y \mid Z = z),$$

where the summation is over all y in the range of Y.

Lemma. *For any random variables X and Y,*

$$\mathbf{E}[X] = \sum_{y} \Pr(Y = y)\mathbf{E}[X \mid Y = y],$$

where the sum is over all values in the range of Y and all of the expectations exist.

Lemma. *For any finite collection of discrete random variables $X_1, X_2, \ldots, X_n$ with finite expectations and for any random variable Y,*

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} \mathbf{E}[X_i \mid Y = y].$$

Definition. *The expression $\mathbf{E}[X \mid Y]$ is a random variable $f(Z)$ that takes on the value $\mathbf{E}[Y \mid Z = z]$ when $Z = z$.*

Theorem.

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y \mid Z]].$$

Definition. *We perform a sequence of independent trials until the first success, where each trial succeeds with probability p. A geometric random variable X with parameter p is given by the following probability distribution on $n = 1, 2, \ldots,$:*

$$\Pr(X = n) = (1-p)^{n-1}p.$$

That is, for the geometric random variable X to equal n, there must be $n-1$ failures, followed by a success.

Lemma. *For a geometric random variable X with parameter p and for $n > 0$,*

$$\Pr(X = n + k \mid X > k) = \Pr(X = n).$$

Lemma. *Let X be a discrete random variable that takes on only non-negative integer values. Then*

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \Pr(X \geq i).$$

Lemma. *For a geometric random variable X with parameter p,*

$$\mathbf{E}[X] = \frac{1}{p}.$$

Lemma. *The harmonic number $H(n) = \sum_{i=1}^{n} 1/i$ satisfies $H(n) = \ln n + \Theta(1)$.*

3. Moments and Deviations

Theorem (Markov's Inequality)**.** *Let X be a random variable that assumes only non-negative values. Then, for all $a > 0$,*

$$\Pr(X \geq a) \leq \frac{\mathbf{E}[X]}{a}.$$

Definition. *The kth moment of a random variable X is $\mathbf{E}[X^k]$.*

Definition. *The* variance *of a random variable X is defined as*

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

The standard deviation *of a random variable X is*

$$\sigma[X] = \sqrt{\mathbf{Var}[X]}.$$

Lemma. *For a Bernoulli random variable with success probability p,*

$$\mathbf{Var}[X] = p(1-p).$$

Definition. *The* covariance *of two random variables X and Y is*

$$\mathbf{Cov}(X, y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

Theorem. *For any two random variables X and Y,*

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X, Y).$$

Theorem. *If X and Y are two independent random variables, then*

$$\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Corollary. *If X and Y are independent random variables, then*

$$\mathbf{Cov}(X, Y) = 0$$

and

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y].$$

Theorem. *Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables. Then*

$$\mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i].$$

Lemma. *For a binomial random variable X with parameters n and p,*

$$\mathbf{Var}[X] = np(1-p).$$

Theorem (Chebyshev's Inequality)**.** *For any $a > 0$,*

$$\Pr(|X - \mathbf{E}[X]| \geq a) \leq \frac{\mathbf{Var}[X]}{a^2}.$$

Corollary. *For any $t > 1$,*

$$\Pr(|X - \mathbf{E}[X]| \geq t \cdot \sigma[X]) \leq \frac{1}{t^2} \text{ and}$$
$$\Pr(|X - \mathbf{E}[X]| \geq t \cdot \mathbf{E}[X]) \leq \frac{\mathbf{Var}[X]}{t^2(\mathbf{E}[X])^2}.$$

Lemma. *For a geometric random variable X with parameter p,*

$$\mathbf{Var}[X] = (1-p)/p^2.$$

Definition. *The X be a random variable. The* median *of X is defined to be any value m such that*

$$\Pr(X \leq m) \geq \frac{1}{2} \quad \text{and} \quad \Pr(X \geq m) \geq \frac{1}{2}.$$

4. Chernoff and Hoeffding Bounds

Definition. *The* moment generating function *of a random variable X is*

$$M_X(t) = \mathbf{E}[e^{tX}].$$

Theorem. *Let X be a random variable with moment generating function $M_X(t)$. Under the assumption that exchanging the expectation and differentiation operands is legitimate, for all $n > 1$ we have*

$$\mathbf{E}[X^n] = M_X^{(n)}(0),$$

*where $M_X^{(n)}(0)$ is the n*th derivative *of $M_X(t)$ evaluated at $t = 0$.*

Theorem. *Let X and Y be two random variables. If*

$$M_X(t) = M_Y(t)$$

for all $t \in (-\delta, \delta)$ for some $\delta > 0$, then X and Y have the same distribution.

Theorem. *If X and Y are independent random variables, then*

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Definition. *A sum of independent 0-1 random variables are known as* Poisson trials*. The distributions of the random variables in Poisson trials are not necessarily identical.* Bernoulli trials *are a special case of Poisson trials where the independent 0-1 random variables have the same distribution; in other words, all trials are Poisson trials that take on the value 1 with the same probability.*

Theorem. *Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. Then the following Chernoff bounds hold:*

(1) for any $\delta > 0$,

$$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu};$$

(2) for $0 < \delta \leq 1$,

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3};$$

(3) for $R \geq 6\mu$,

$$\Pr(X \geq R) \leq e^{-R}.$$

Theorem. *Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. Then, for $0 < \delta < 1$:*

(1)

$$\Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu};$$

(2)

$$\Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}$$

Corollary. *Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. For $0 < \delta < 1$,*

$$\Pr(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\delta^2/3}.$$

Theorem. *Let $X_1, \ldots, X_n$ be independent random variables with*

$$\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_{i=1}^{n} X_i$. For any $a > 0$,

$$\Pr(X \geq a) \leq e^{-a^2/2n}.$$

By symmetry we also have

$$\Pr(X \leq -a) \leq e^{-a^2/2n}.$$

Corollary. *Let $X_1, \ldots X_n$ be independent random variables with*

$$\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_{i=1}^{n} X_i$. Then, for any $a > 0$,

$$\Pr(|X| \geq a) \leq 2e^{-a^2/2n}.$$

Corollary. *Let $Y_1, \ldots, Y_n$ be independent random variables with*

$$\Pr(Y_i = 1) = \Pr(Y_i = 0) = \frac{1}{2}.$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = \mathbf{E}[Y] = n/2$.

(1) For any $a > 0$,

$$\Pr(Y \geq \mu + a) \leq e^{-2a^2/n}.$$

(2) For any $\delta > 0$,

$$\Pr(Y \geq (1+\delta)\mu) \leq e^{-\delta^2\mu}.$$

Corollary. *Let $Y_1, \ldots, Y_n$ be independent random variables with*

$$\Pr(Y_i = 1) = \Pr(Y_i = 0) = \frac{1}{2}.$$

Let $Y = \sum_{i=1}^{n} Y_i$ and $\mu = \mathbf{E}[Y] = n/2$.

(1) For any $0 < a < \mu$,

$$\Pr(Y \leq \mu - a) \leq e^{-2a^2/n}.$$

(2) For any $0 < \delta < 1$,

$$\Pr(Y \leq (1-\delta)\mu) \leq e^{-\delta^2 \mu}.$$

Theorem (Hoeffding Bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that for all $1 \leq i \leq n$, $\mathbf{E}[X_i] = \mu$ and $\Pr(a \leq X_i \leq b) = 1$. Then*

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Lemma (Hoeffding's Lemma)**.** *Let X be a random variable such that $\Pr(X \in [a,b]) = 1$ and $\mathbf{E}[X] = 0$. Then for every $\lambda > 0$,*

$$\mathbf{E}[e^{\lambda X}] \leq e^{\lambda^2 (b-a)^2/8}.$$

Theorem. *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbf{E}[X_i] = \mu_i$ and $\Pr(a_i \leq X_i \leq b_i) = 1$ for constants a_i and b_i. Then*

$$\Pr\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \geq \epsilon\right) \leq 2e^{-2\epsilon^2 / \sum_{i=1}^{n} (b_i - a_i)^2}.$$

13. Martingales

Definition. *A sequence of random variables $Z_0, Z_1, \ldots$ is a* martingale *with respect to the sequence $X_0, X_1, \ldots$ if, for all $n \geq 0$, the following conditions hold:*

- *Z_n is a function of $X_0, X_1, \ldots, X_n$;*
- *$\mathbf{E}[|Z_n|] < \infty$;*
- *$\mathbf{E}[Z_{n+1} \mid X_0, \ldots, X_n] = Z_n$.*

A sequence of random variables $Z_n, Z_1, \ldots$ is called a martingale when it is a martingale with respect to itself. That is, $\mathbf{E}[|Z_n|] < \infty$, and $\mathbf{E}[Z_{n+1} \mid Z_0, \ldots, Z_n] = Z_n$.

Definition. *A* Doob martingale *refers to a martingale constructed using the following general approach. Let $X_0, X_1, \ldots, X_n$ be a sequence of random variables, and let Y be a random variable with $\mathbf{E}[|Y|] < \infty$. (Generally, Y will depend on $X_0, \ldots, X_n$.) Then*

$$Z_i = \mathbf{E}[Y \mid X_0, \ldots, X_i], \quad i = 0, 1, \ldots, n,$$

gives a martingale with respect to $X_0, X_1, \ldots, X_n$.

Lemma. *If the sequence $Z_0, Z_1, \ldots, Z_n$ is a martingale with respect to $X_0, X_1, \ldots, X_n$, then*

$$\mathbf{E}[Z_n] = \mathbf{E}[Z_0].$$

Definition. *A nonnegative integer-valued random variable T is a* stopping time *for the sequence $\{Z_i, i \geq 0\}$ if the probability of the event $T = n$ is independent of variables $\{Z_{n+j} \mid Z_1, \ldots Z_n, j \geq 1\}$ (i.e., the variables $Z_{n+1}, Z_{n+2}, \ldots$ conditioned on the values $Z_1, \ldots, Z_n$).*

Theorem (Martingale Stopping Theorem)**.** *If $Z_0, Z_1, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$ and if T is a stopping time for $X_1, X_2, \ldots$, then*

$$\mathbf{E}[Z_T] = \mathbf{E}[Z_0]$$

whenever one of the following holds:

- *the Z_i are bounded, so there is a constant c such that, for all i, $|Z_i| \leq c$;*
- *T is bounded;*
- *$\mathbf{E}[T] < \infty$, and there is a constant c such that $\mathbf{E}[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] < c$.*

Theorem (Wald's Equation)**.** *Let $X_1, X_2, \ldots$ be nonnegative, independent, indentically distributed random variables with distribution X. Let T be a stopping time for this sequence. If T and X have bounded expectation, then*

$$\mathbf{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbf{E}[T] \cdot \mathbf{E}[X].$$

Definition. *Let $Z_0, Z_1, \ldots$ be a sequence of independent random variables. A nonnegative, integer-valued random variable T is a stopping time for the sequence if the event $T = n$ is independent of $Z_{n+1}, Z_{n+2}, \ldots$.*

Theorem (Azuma-Hoeffding Inequality)**.** *Let $X_0, \ldots, X_n$ be a martingale such that*

$$|X_k - X_{k-1}| \leq c_k.$$

Then, for all $t \geq 1$ and $\lambda > 0$,

$$\Pr(|X_t - X_0| \geq \lambda) \leq 2e^{-\lambda^2/2\sum_{k=1}^{t} c_k^2}.$$

Corollary. *Let $X_0, X_1, \ldots$ be a martingale such that, for all $k \geq 1$,*

$$|X_k - X_{k-1}| \leq c.$$

Then, for all $t \geq 0$ and $\lambda > 0$,

$$\Pr(|X_t - X_0| \geq \lambda c\sqrt{t}) \leq 2e^{-\lambda^2/2}.$$

Theorem (Azuma-Hoeffding Inequality)**.** *Let $X_0, \ldots, X_n$ be a martingale such that*

$$B_k \leq X_k - X_{k-1} \leq B_k + d_k$$

for some constants d_k and for some random variables B_k that may be functions of $X_0, X_1, \ldots, X_{k-1}$. Then, for all $t \geq 0$ and any $\lambda > 0$,

$$\Pr(|X_t - X_0| \geq \lambda) \leq 2e^{-2\lambda^2/(\sum_{k=1}^{t} d_k^2)}.$$

Definition. *A function*

$$f(\bar{X}) = f(X_1, X_2, \ldots, X_n)$$

satisfies a Lipschitz condition *with bound c if, for any i and for any set of values $x_1, \ldots, x_n$ and y_i,*

$$|f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, x_2, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)| \leq c.$$

Theorem (McDiarmid's Inequality)**.** *Let f be a function on n variables that satisfies the above Lipschitz condition with bound c. Let $X_1, \ldots, X_n$ be independent random variables such that $f(X_1, \ldots, X_n)$ is in the domain of f. Then*

$$\Pr(|f(X_1, \ldots, X_n) - \mathbf{E}[f(X_1, \ldots, X_n)]| \geq \lambda) \leq 2e^{-2\lambda^2/(nc^2)}.$$