CSCI 2570 Introduction to Nanocomputing

Probability Theory

John E Savage

The Role of Probability

- The manufacture of devices with nanometer-scale dimensions will necessarily introduce randomness into these devices.
- Some device dimensions are so small that their position cannot be accurately controlled
- For this reason, probability theory will play a central role in this area

Sample Spaces

- Probabilities estimate the frequency of outcomes of random experiments.

- Outcomes can be from a finite or countable **sample space** (set) Ω of **events** or be tuples drawn over reals ***R***.
  - Coin toss: Ω = {H,T}
  - Packets to a URL per day: Ω = ***N*** (positive integers)
  - Rain in cms/month in Prov.: Ω = ***R*** (reals)
  - Rain and sunshine/month: Ω = $\boldsymbol{R}^2$

Probability Space

- **Sample space**: all possible outcomes
- **Events**: A family ***F*** of subsets of sample space Ω.
  - E.g. $\Omega = \{H,T\}^3$, F_0 = {TTT, HHT, HTH, THH} (Even no. Hs). F_1 = {HTT, THT, TTH, HHH} (Odd no. Hs).
- Events are **mutually exclusive** if they are disjoint. E.g. F_0 and F_1 above.
- A **probability distribution** is a function $p : \Omega \mapsto \mathcal{R}$
- The probability distribution assigns a **probability** $0 \le P(E) \le 1$ to each event E.

Properties of Probability Function

- For any event E in Ω, $0 \leq P(E) \leq 1$.
- $P(\Omega) = 1$
- For any finite or countably infinite sequence of disjoint events $E_1, E_2, \ldots$

$$Pr(\cup_{i \geq 1} E_i) = \sum_{i \geq 1} P(E_i)$$

Probability Distributions

- If $\Omega = \boldsymbol{R}^n$, **probability density** $p(x_1,\ldots x_n)$ can be integrated over a volume to give a probability. E.g. $A = \{2 \leq x \leq 3.5\}$, $B = \{y \leq 15\}$

$$P(A) = \int_2^{3.5} \int_{-\infty}^{\infty} p(x,y)\, dx\, dy$$

$$P(A,B) = \int_2^{3.5} \int_{-\infty}^{15} p(x,y)\, dx\, dy$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x,y)\, dx\, dy = 1$$

Sets of Events

- **Joint probability** P(A∩B) = $\sum_{e \in A \cap B} p(e)$
  - Notation: P(A,B) = P(A∩B)

- **Probability of a** union P(A∪B) = $\sum_{e \in A \cup B} p(e)$

  P(A∪B) = P(A) + P(B) – P(A∩B)

- **Complement** of event A: $\overline{A}$ = Ω–A.P(A∪$\overline{A}$)=1

Probabilities of Events

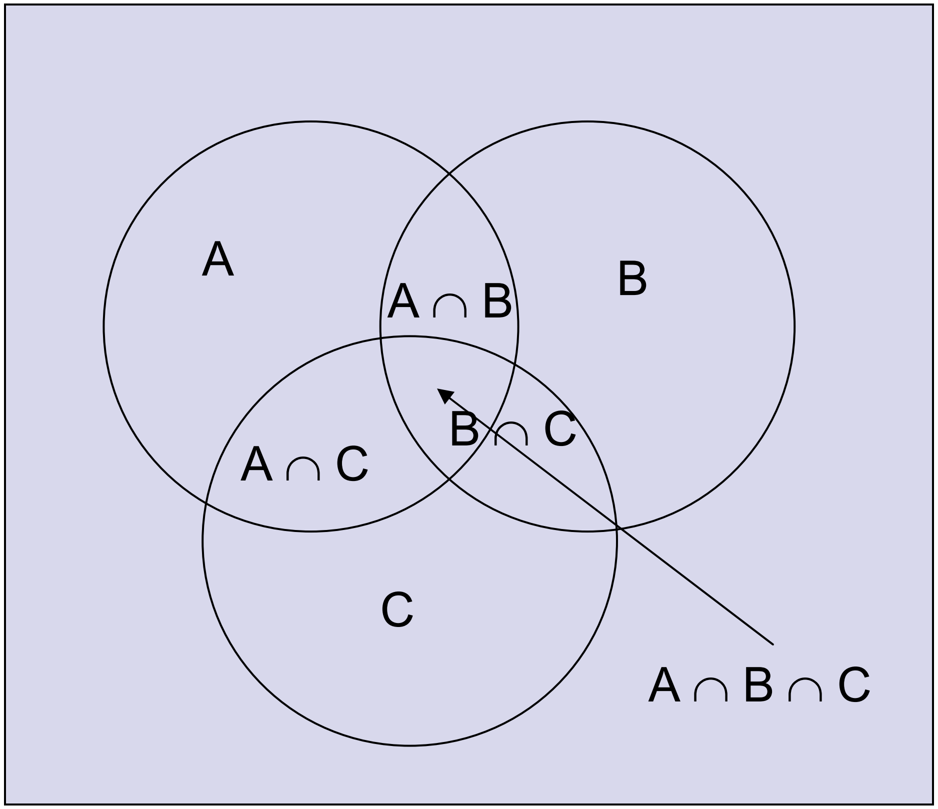
- If events A and B are **mutually exclusive**
  - $P(A \cap B) = 0$
  - $P(A \cup B) = P(A) + P(B)$
- **Conditional probability** of A given B, $P(A/B) = P(A,B)/P(B)$ or $P(A,B) = P(A/B)P(B)$.
- Events A and B are statistically **independent** if $P(A/B) = P(A)$, i.e., $P(A,B) = P(A)P(B)$

Marginal Probability

- Given a sample space $\Omega = K^2$ containing pairs of events A_i, B_j over K, the **marginal probability** is $P(A) = \sum_j P(A, B_j)$, where B_j are mutually exclusive.

Principle of Exclusion/Inclusion

- Let |A| = size of A
- $|A \cup B| = |A|+|B| - |A \cap B|$
- $|A \cup B \cup C| = |A|+|B|+|C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$



Principle of Inclusion/Exclusion

$$Pr(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} Pr(E_i) - \sum_{i<j} Pr(E_i \cap E_j) + \sum_{i<j<k} Pr(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n+1} \sum_{i_1<i_2<\cdots<i_n} Pr(\bigcap_{i=1}^{n} E_i)$$

Proof Use induction. Assume true for *n-1* sets.

Let $F_i = E_i$ for $1 \le i \le n-2$ and let $F_{n-1} = E_{n-1} \cup E_n$ and apply $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Application of Inclusion/Exclusion

- For l odd, $(-1)^{l+1} = 1$

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

- For l even , $(-1)^{l+1} = -1$

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

Special Application of Inclusion/Exclusion

$$\sum_{i=1}^{n} Pr(E_i) - \sum_{i<j} Pr(E_i \cap E_j) \leq Pr(\bigcup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} Pr(E_i)$$

Event Product Spaces

- Important sample spaces consists of Cartesian products of spaces
  - $\Omega = \{(H,H), (H,T), (T,H), (T,T)\} = \{H,T\}^2$
  - $\Omega = A^n = \{e_1, e_2, \ldots, e_n\}$, e_i in A.
  - $P_{1,2}(H,H)$ = prob. of event (H,H).
  - E.g. P(H,H) =.04, P(H,T)=P(T,H) =.16,P(T,T) =.64

- They can model occurrences over time or space or both

Event Product Spaces

- Given events A and B with joint probability P(A,B), P(A) is the marginal probability of A.
- E.g.
  - $P_1(H) = P_{1,2}(H,H) + P_{1,2}(H,T) = .04 + .16 = .20$
  - $P_1(T) = P_{1,2}(T,H) + P_{1,2}(T,T) = .16 + .64 = .80$
- Consider events H and T on successive trials that are independent.
  - E.g. $P_{1,2}(H,T) = P_1(H)\, P_2(T) = .2 \times .8 = .16$

Product Events

- Events are **identically distributed** if they have the same probability distribution.
  - Outcomes in a pair of H,T trials are i.d.
  - $P_1 = P_2$, that is, $P_1(e) = P_2(e)$ for all e in {H,T}
- Events are **independent and identically distributed** (**i.i.d.**) if they are statistically independent and identically distributed.

Random Variables

- A **random variable** *v* is a function $v : \Omega \mapsto \mathcal{R}$
  - E.g. Ω = {H,T}, v(H) = 1, v(T) = 0
- **Expectation (average value)** of a r.v. *v* is
  $$E(v) = \overline{x} = \sum_{e \in \Omega} v(e)p(e)$$
  - E.g. $\overline{x} = 1 \times .2 + 0 \times .8 = .2$
- Expectation of sum is sum of expectations
  $$E(x_1 + \cdots + x_n) = E(x_1) + \cdots + E(x_n)$$

Geometric Random Variable

$$Pr(n) = (1-p)^{n-1}p \text{ for } 0 \leq n$$

$$\overline{n} = \sum n(1-p)^{n-1}p = p\frac{d\left(\sum z^n\right)}{d\,z}\,\Big|_{z=1-p}$$

$$\overline{n} = p\frac{d}{d\,z}(1-z)^{-1}\,\Big|_{z=1-p} = 1/p$$

Moments of Random Variables

- **Second moment** of a r.v. $E(v^2) = \sum_e v^2(e)p(e)$
- ***k*th moment** or a r.v. $E(v^k) = \sum_e v^k(e)p(e)$
- **Variance**

$$Var(v) = \sigma^2 = E((v - E(v))^2) = E(v^2) - E^2(v)$$

- **Standard deviation** $\sigma = \sqrt{Var(v)}$

Examples of Probability Distributions

- **Uniform**: $P(k) = 1/n$ for $1 \le k \le n$

- **Binomial**: n i.i.d. trials, $\Omega = \{H,T\}^n$, $P(H) = \alpha$ and $P(T) = \beta = 1 - \alpha$. $P(k) = Pr(k\ H\text{'s occur})$

$$P(k) = \binom{n}{k} \alpha^k \beta^{n-k}, \quad 0 \le k \le n$$

- **Poisson**: $P_\nu(n) = \frac{\nu^n e^{-\nu}}{n!}, \quad 0 \le n < \infty$
  - Is limit of binomial when $\nu = \alpha n$ and n large.

Means and Variances of Probability Distributions

- **Uniform**: $\overline{x} = \sum_{k=1}^{n} k/n = (n+1)/2$

  $$\overline{x^2} = \sum_{k=1}^{n} k^2/n = (n+1)(n+1/2)/3$$

- **Binomial**: $\overline{x} = n\alpha$

  $$\overline{x^2} = \sigma^2 + E^2(x),\ \sigma = \sqrt{n\alpha\beta}$$

- **Poisson**: $\overline{x} = \nu$

  $$\overline{x^2} = \sigma^2 + E^2(x),\ \sigma = \sqrt{\nu}$$

Markov's Inequality

- Let X be a **positive** r.v., $Pr(X \geq a) \leq \frac{E(X)}{a}$

Proof Because $1 \leq x/a$ when $x \geq a$

$$
\begin{aligned}
Pr(x \geq a) &= \sum_{x \geq a} p(x) \\
&\leq \sum_{x \geq a} p(x)(x/a) \\
&\leq \sum_{x} p(x)(x/a) \\
&= \frac{E(x)}{a}
\end{aligned}
$$

Chebyshev's Inequality

- Let X be a r.v. $Pr(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$

Proof Note $1 \leq ((x - \overline{x})/a)^2$ when $|x - \overline{x}| \geq a$

Let $A = \{x \text{ such that } |x - E(x)| \geq a\}$

$$
\begin{aligned}
Pr(|X - E(X)| \geq a) &= \sum_{x \in A} p(x) \\
&\leq \sum_x p(x) \frac{(x - \overline{x})^2}{a^2} \\
&= \frac{Var(x)}{a^2}
\end{aligned}
$$

Moment Generating Function

- $g(t) = \overline{e^{tx}}$ is a function that can be used to compute moments and Chernoff bounds on tails of probabilities, i.e. $P(x \geq X)$

$$\overline{x} = \frac{d\, g(t)}{d\, t} |_{t=0} \qquad \overline{x^2} = \frac{d^2\, g(t)}{d\, t^2} |_{t=0}$$

$$\overline{x^k} = \frac{d^k\, g(t)}{d\, t^k} |_{t=0}$$

Moment Generating Functions

- **Uniform**:

$$g_U(t) = \sum_{k=1}^{n} e^{tk}\,\frac{1}{n} = \frac{1}{n}\frac{e^{t(n+1)} - e^t}{e^t - 1}$$

- **Binomial**:

$$g_B(t) = \sum_{k=0}^{n} e^{tk}\binom{n}{k}\alpha^k \beta^{n-k} = \left(e^t \alpha + \beta\right)^n$$

- **Poisson**:

$$g_B(t) = \sum_{n=0}^{\infty} e^{tn}\,\frac{\nu^n e^{-\nu}}{n!} = \sum_{n=0}^{\infty} \frac{(\nu e^t)^n e^{-\nu}}{n!} = e^{\nu(e^t - 1)}$$

Chernoff Bound

- Let X be a r.v. $Pr(X \geq a) \leq e^{-ta} g(t)$ for $t > 0$.

Proof Because $e^{t(x-a)} \geq 1$ when $x \geq a$ & $t \geq 0$

$$\begin{aligned} Pr(X \geq a) &= \sum_{x \geq a} p(x) \\ &\leq \sum_{x} p(x) e^{t(x-a)} \\ &= \frac{g(t)}{e^{ta}} \end{aligned}$$

Bounding Tails of a Binomial

- $E(x) = n\alpha,\ Var(x) = \sqrt{n\alpha\beta}$

  $g(t) = \sum_{k=0}^{n} \binom{n}{k} \alpha^k \beta^{n-1} e^{tk} = (\alpha e^t + \beta)^n$

- Markov $Pr(X \geq a) \leq \frac{E(X)}{a} = \frac{n\alpha}{a}$

- Chebyshev $Pr(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2} = \frac{n\alpha\beta}{a^2}$

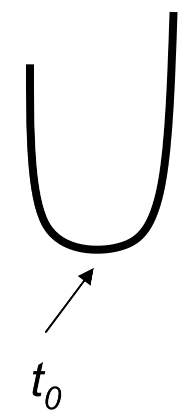
- Chernoff $Pr(X \geq a) \leq e^{-ta} g(t) = e^{-ta} (\alpha e^t + \beta)^n$

Chernoff Bound on Binomial Distribution

- $Pr(X \geq a) \leq e^{-ta} g(t) = e^{-ta}(\alpha e^t + \beta)^n$
  - Choose $t = t_0$ to minimize bound
  - Note that $e^{-ta} g(t) = E(e^{t(z-a)})$ is convex because its second derivative is positive.
  - Thus, at t_0 the first derivative is zero.
  - That is $t_0 = \ln\left(\frac{a\beta}{n}\right) - \ln\left(\left(1 - \frac{a}{n}\right)\alpha\right)$ and

    $Pr(X \geq a) \leq e^{\theta(n,\alpha)}$ where

    $\theta(n, \alpha) = n(\rho \ln \alpha + (1-\rho) \ln \beta + H(\rho))$
  - Here $\rho = a/n$ and $H(y) = -y \ln y - (1-y) \ln(1-y)$



Comparison of Bounds

- *n=100, α=.5, β=.5, a=70, E(x)=50, Var(x) = 5*
- **Markov**: $Pr(X \geq 70) \leq \frac{E(X)}{70} = \frac{50}{70} = .714$
- **Chebyshev**:$Pr(|X - 50| \geq 20) \leq \frac{25}{400} = .0625$
  implies $Pr(X \geq 70) \leq 2 \times \frac{25}{400} = .125$
- **Chernoff**: $\rho = .7 \text{ and } H(\rho) = .61086$
  $\theta(\rho, \alpha) = n(\rho \ln \alpha + (1-\rho) \ln \beta + H(\rho)) = -8.228$
  implies $Pr(X \geq 70) \leq e^{\theta(\rho,\alpha)} = .000267$
- **Exact**: $Pr(X \geq 70) = .00003$

Birthday Problem

- Each person equally likely to have day x as birthday, $1 \leq x \leq 365$
- In a group of n persons, what is probability P_B that at least two have same birthday?
  - $1-P_B = 365(365-1)\ldots(365-n+1)/365^n$
  - $P_B \approx .5$ when $n \approx 23$!

Balls in Bins

- *m* balls thrown into *n* bins independently and uniformly at random
- How large should *m* be to ensure that all bins contain at least one ball with prob. $\geq 1-\varepsilon$?
- **Coupon collector problem**:
  - C coupon types
  - Each box equally likely to contain any coupon type
  - How many boxes should be purchased to collect all coupons with probability at least $1-\varepsilon$?

Coupon Collector Problem

- C coupons, one per box with probability $1/C$ in a box
- What is $E(X)$, X = no. boxes to collect all coupons?
- $X = x_1+\ldots+x_C$, x_i = no. boxes until ith coupon is collected. Prob. of a new coupon: $p_i = 1-(i-1)/C$
- x_i is geometric r.v. with $Pr(x_i = n) = (1-p_i)^{n-1}p_i$
  - $E(x_i) = 1/p_i = C/(C-i+1)$
- $E(X)=E(x_1)+\ldots+E(x_C) = \sum_{i=1}^{C} \frac{C}{C-i+1} = C\sum_{j=1}^{C}\frac{1}{j} \approx C \ln C$

Coupon Collector Problem with Failures

In this model the probability that a coupon is not collected is 1-p_s. The probability that a specific coupon is collected is p_s/C.

Theorem Let T = no. trials to ensure all C coupons collected with probability = 1-ε in coupon collector problem with failures satisfies

$$\frac{C}{p_s(1+p_s/C)} \ln\left(\frac{C}{\epsilon(1+\epsilon)}\right) \leq T \leq \frac{C}{p_s} \ln\left(\frac{C}{\epsilon}\right)$$

Special Application of Inclusion/Exclusion

$$\sum_{i=1}^{n} Pr(E_i) - \sum_{i<j} Pr(E_i \cap E_j) \leq Pr(\bigcup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} Pr(E_i)$$

Coupon Collection with Failures

Proof Let E_i be event *i*th coupon not collected after T trials. $P(E_i) = (1 - p_s/C)^T$ Also

$$P(E_i \cap E_j) = (1 - p_i - p_j)^T = (1 - 2p_s/C)^T$$

The goal is to find T so that $Pr(\bigcup_{i=1}^n E_i) = \epsilon$

Using Inclusion/Exclusion & $(1 - 2x) \leq (1 - x)^2$

$$Pr(\bigcup_{i=1}^n E_i) \leq \sum_{i=1}^n Pr(E_i) = C(1 - p_s/C)^T$$

$$\sum_{i=1}^n Pr(E_i) - \sum_{i<j} Pr(E_i \cap E_j) \leq Pr(\bigcup_{i=1}^n E_i)$$

$$C(1 - p_s/C)^T - \frac{C^2}{2}(1 - p_s/C)^{2T} \leq$$

Coupon Collection with Failures

Then

$$C\,(1 - p_s/C)^T \left[1 - \frac{C}{2}(1 - p_s/C)^T\right] \leq \epsilon \leq C\,(1 - p_s/C)^T$$

Equivalently $z(1 - z/2) \leq \epsilon \leq z$ for $z = C\,(1 - p_s/C)^T$

but this implies

$$\epsilon \leq z \leq \epsilon(1 + \epsilon) \text{ if } \epsilon \leq \sqrt{2} - 1 = .414214$$

Using $e^{-x(1+x)} \leq 1 - x \leq e^{-x}$ when $x = p_s/C \leq .5$ or $C \geq 2$

gives the desired result.

Conclusion

- Methods of bounding tails of probability distributions can be very useful.