Random Variables

- A **random variable** X on a sample space Ω is a function that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.
- Discrete r.v.'s take values in a range that is finite or countably infinite.
- Since a r.v. is defined on a probability space, we can calculate these probabilities given the probabilities of the sample points.
- Let a be any number in the range of a r.v. X. Then the set $\{\omega \in \Omega : X(\omega) = a\}$ is an event in the sample space.
- The **distribution** of a discrete r.v. X is the collection of values $\{(a, Pr[X = a]) : a \in \mathscr{A}\}$
- The collection of events $X = a, a \in \mathscr{A}$, satisfy two important properties:
  - any two events $X = a_1$ and $X = a_2$ with $a_1 \neq a_2$ are disjoint
  - the union of all these events is equal to the entire sample space Ω

Binomial Distribution

- $Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$ where $i = 0, 1, \ldots, n$
  $Pr[X \geq n] = \sum_{i=n}^{n+k} \binom{n+k}{i} (1-p)^i p^{n+k-i}$
- The **expectation** (or *mean* or *average*) of a discrete random variable X is defined as $E(X) = \sum_{a \in \mathscr{A}} a \times Pr[X = a]$

Linearity of Expectation

- $E(X + Y) = E(X) + E(Y)$
- $E(cX) = cE(X)$

Variance

- "Spread" of distances from the mean.
- $Var(X) = E((X - \mu)^2) = E(X^2) - \mu^2$
- Standard deviation, $\sigma = \sqrt{Var(X)}$
- $Var(cX) = c^2 Var(X)$
- $Var(X + Y) = Var(X) + Var(Y)$ (**independent**)
- $E(XY) = E(X)E(Y)$ (**independent**)
- $Var(X + Y) = Var(X) + Var(Y) + 2(E(XY) - E(X)E(Y))$

If X is uniform random variable $\{1 \ldots n\}$ with prob $\frac{1}{n}$:

- $E(X) = \frac{n+1}{2}$
- $\text{Var}(X) = \frac{n^2-1}{12}$
- $\sigma(X) = \sqrt{\frac{b^2-1}{12}}$
- $Var(X_i) = E(X_i^2) - E(X_i)^2$

Continuous Probability

- integral = 1
- $E[X] = \int_{-\infty}^{\infty} XP(X)dx = \sum XP(X)$
- $P[X_0 < X < X_1] = \int_{X_0}^{X_1} XP(X)dx$
- $Var(x) = E[(X - \mu)^2] = E[X^2] - \mu^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

Central Limit Theorem

- $A_n = \frac{\sqrt{n}(A_n - \mu)}{\sigma} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$
- $\sigma_\mu = \frac{\sigma}{\sqrt{n}}$. plug in to normal pdf and intergate to get intervals.

Geometric Distribution

- Flip n coins, stop after first heads. P($heads$) = p.
- $P[X = i] = (1-p)^i p$
- $E(X) = \frac{1}{p}$
- $Var(X) = \frac{1-p}{p^2}$
- $Pr[X \geq i] = (1-p)^{i-1}$ for $i = 1, 2, \ldots$

Poisson Distribution

- $P[X] = 0$ iff $x < 0$ AND $P[x] = \lambda e^{-\lambda x}$ for $x \geq 0$
- $E[X] = \lambda$
- $P[X = i] = \frac{\lambda^i e^{-\lambda}}{i!}$
- $Var(x) = \lambda$

Exponential Distribution

- $f(x) = \lambda e^{-\lambda x}$ for $x > 0$, otherwise 0.
- $E(x) = \frac{1}{\lambda}$
- $Var(x) = \frac{1}{\lambda^2}$
- $P[X \geq C] = \int_C^{\infty} \lambda e^{-\lambda x} dx$

Normal Distribution

- Parameters σ and μ, centered at $x = \mu$, standard deviation is σ
- standard normal: $\mu = 0$ and $\sigma = 1$
- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$
- $E(x) = \mu$
- $Var(x) = \sigma^2$
- $P[X \geq C] = \int_C^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$ (just integrate the interval)
- Note: 68% of the data lies within 1σ from the mean, 95% within 2σ, and 99.7% within 3σ

Probability Density Function

- $f(x) = 0$ for $x < 0$ and $x > \ell$
- $f(x) = \frac{1}{\ell}$ for $0 \leq x \leq \ell$
- $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \frac{\ell}{2}$
- $Var(X) = \frac{\ell^2}{12}$

Formulas

- Chebyshev (random var X with $E(X) = \mu$, $\alpha > 0$):
  $Pr[|X - \mu| \geq \alpha] \leq \frac{Var(X)}{\alpha^2}$
- Chebyshev part 2: (random var X with $E(X) = \mu$, $\sigma = \sqrt{Var(X)}$):
  $Pr[|X - \mu| \geq \beta\sigma] \leq \frac{1}{\beta^2}$
- Markov's inequality ($X > 0$, $E(X) = \mu$, $\alpha > 0$):
  $Pr[X \geq \alpha] \leq \frac{E(X)}{\alpha}$
- Joint Density function:
  $Pr[a \leq X \leq b, c \leq Y \leq d] = \int_c^d \int_a^b f(x, y)dxdy$
- probability per unit area:
  $Pr[x \leq X \leq x + \delta, y \leq Y \leq y + \delta] = \int_y^{y+\delta} \int_x^{x+\delta} f(u, v)dudv \approx \delta^2 f(x, y)$
- Independent variables x, y, then $f(x, y) = f(x)f(y)$

Simpsons paradox: A paradox in which a trend that appears in different groups of data disappears when these groups are combined.
Cardinality: In order to determine whether two sets have the same cardinality, we need to demonstrate a *bijection* f between the two sets.
We say that a set S is **countable** if there is a bijection between S and $\mathbb{N}$ or some subset of $\mathbb{N}$. Thus any finite set S is countable (since there is a bijection between S and the subset $\{0, 1, 2, \ldots, m-1\}$, where $m = \mid S \mid$ is the size of S).
If there is a one-to-one function $f : A \to B$, then the cardinality of A is less than or equal to that of B. Now to show that the cardinality of A and B are the same we can show that $\mid A \mid \leq \mid B \mid$ and $\mid B \mid \leq \mid A \mid$.
Cantors Diagonalization: $\mathscr{P}(S) = \{T : T \subseteq S\}$.

Sets

- $\mathbb{Z}$ - integers (countable. f(x)=2x, f(-x)=-2x+1)
- $\mathbb{N}$ - natural (countable)
- $\mathbb{Q}$ - rational (countable - N/N. also spiral method)
- $\mathbb{R}$ = real (uncountable - always can find avg(a,b))
- Union of countable and uncountable must yield uncountable set.

The Halting Problem:
Given the description of a program and its input, we would like to know if the program ever halts when it is executed on the given input.

$$TestHalt(P, I) = \begin{cases} \text{"yes"} & \text{if program P halts on input I} \\ \text{"no"} & \text{if program P loops on input I} \end{cases}$$

Proof: Define the program
Turing(P)
if TestHalt(P,P) = "yes" then loop forever
else halt
So if the program P when given P as input halts, then Turing loops forever; otherwise, Turing halts.
TestHalt cannot exist, so it is impossible for a program to check if any general program halts.