Author: Ethan Schreiber - ethan@cs.ucla.edu

Axioms: $\mathbf{P}(A) \geq 0$	disjoint $A_i \implies \bigcup_{i=1}^n A_i = \sum_{i=1}^n A_i$	$P(\Omega) = 1$

Thms: $A \subset B \implies P(A) \leq P(B)$ $\mathbf{P}(A \cup B) \leq P(A) + P(B)$ $\mathbf{P}(A) = \mathbf{P}(A \cap B) \cup \mathbf{P}(A \cap B^C)$ (disjoint)	$\mathbf{P}(A \cup B) = P(A) + P(B) - P(A \cap B)$ $\mathbf{P}(A \cup B \cup C) = P(A) + P(A^C \cap B) + P(A^C \cap B^C \cap C)$
$A \subset B \implies [\mathbf{P}(A \cap B) = \mathbf{P}(A), \quad \mathbf{P}(A \cup B) = \mathbf{P}(B), \quad \mathbf{P}(B - A) = \mathbf{P}(B) - \mathbf{P}(A)]$	

Inclusion-Exclusion: $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$
$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(B \cap C) - \mathbf{P}(A \cap C) + \mathbf{P}(A \cap B \cap C)$

Tot Prob: $\mathbf{P}(B) = \mathbf{P}(A_1 \cap B) + ... + \mathbf{P}(A_n \cap B) = \mathbf{P}(A_1)\mathbf{P}(B|A_1) + ... + \mathbf{P}(A_n)\mathbf{P}(B|A_n)$
Bayes: $\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B|A_i)\sum_{i=1}^k \mathbf{P}(A_i)}$

Counting	Order	No Order
Replacement	n^r	
No Replacement	$\frac{n!}{(n-r)!}$	$\frac{n!}{(n-r!)r!}$

Name	*Definition*	$\mathbf{E}[X]$	$var(X)$	notes
Bernoulli:	$f_X(x) = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases}$	p	$p(1-p)$	
Binomial:	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$p(1-p)$	$\mathbf{P}(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}$
Geometric:	$p_X(k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\mathbf{P}(X > k) = (1 - p^k)$ $\mathbf{P}(X < k) = 1 - (1 - p^k)$ $\mathbf{P}(X = t) = \mathbf{P}(X = k + t - 1 \vert X \geq k)$ $\mathbf{P}(X = t) = \mathbf{P}(X = k + t \vert X > k)$
Exponential:	$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\mathbf{P}(X \geq a) = e^{-\lambda a}$
Poisson	$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ	$\lambda = np$ for large n and small p

Conditional:	$p_{X\vert A}(x) = \mathbf{P}(X = x\vert A) = \frac{P(X=x \cap A)}{\mathbf{P}(A)}$	**Memoryless:**	$PR(X = k + t - 1 \vert X \geq k) = \mathbf{P}(X = t)$

Bloom Filter - False negatives possible but false positives are not. Two hash functions: $h_1(x), h_2(x)$

$$p_Y(y) = \begin{cases} \frac{1}{3}, & \text{if } y = 1 \\ \frac{2}{3}, & \text{if } y = 2 \\ 0 & \text{otherwise} \end{cases} \quad \begin{array}{|ccc} (1,1) & (2,1) & (3,1) \\ (1,2) & (2,2) & (3,2) \\ (1,3) & (2,3) & (3,3) \end{array}$$

$$\mathbf{P}(B_2 = 1) = \mathbf{P}(B_2 = 1|Y = 1)\mathbf{P}(Y = 1) + \mathbf{P}(B_2 = 1|Y = 2)\mathbf{P}(Y = 2) = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{2}{3} = \frac{5}{9}$$

$$\mathbf{P}(B_2 = 1|B_1 = 0) = \frac{\mathbf{P}(B_2 = 1 \cap B_1 = 0)}{\mathbf{P}(B_1 = 0)} = \frac{\mathbf{P}(B_2 = 1 \cap B_1 = 0|Y = 1)P_Y(1) + \mathbf{P}(B_2 = 1 \cap B_1 = 0|Y = 2)\mathbf{P}_Y(2)}{\mathbf{P}(B_1 = 0|Y = 1)\mathbf{P}(Y = 1) + \mathbf{P}(b_1 = 0|Y = 2)\mathbf{P}(Y = 2)}$$
$$= \frac{\frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}}{\frac{2}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3}} = \frac{\frac{6}{27}}{\frac{4}{9}} = \frac{5}{9}$$

lifetime - 3 parallel servers, each server geometric with parameter p. (has prob p of failing per unit time)
Let S_i = lifetime of i^{th} server. Let $T = \max\{S_1, S_2, S_3\}$ be the time until system fails
operational if at least one server operational

$\mathbf{P}(T \leq t) = \mathbf{P}(\max S_1, S_2, S_3 \leq t) \quad = \mathbf{P}(S_1 \leq t) \cdot \mathbf{P}(S_2 \leq t) \cdot \mathbf{P}(S_3 \leq t) \quad = \mathbf{P}(S_1 \leq t)^3 \quad = (1 - \mathbf{P}(S_1 > t))3 \quad = (1 - (1-p)^t)^3$
$\mathbf{P}(T = t) = \mathbf{P}(T \leq t) - \mathbf{P}(T \leq t - 1) = (1 - (1-p)^t)^3 - (1 - (1-p)^{t-1})^3$

operational if at least two servers operational
Let T_i = time from $(i-1)^{th}$ to i^{th} server failure. $T = T_1 + T_2$
$p' = \mathbf{P}(> 1 \text{ fails when 3 are operational}) = 1 - (1-p)^3 \qquad p'' = \mathbf{P}(> 1 \text{ fails when 2 are operational}) = 1 - (1-p)^2$

$$\mathbf{E}[T] = \mathbf{E}[T_1] + \mathbf{E}[T_1] = \frac{1}{p'} + \frac{1}{p''} = \frac{1}{1 - (1-p)^3} + \frac{1}{1 - (1-p)^2}$$

PMF:

$p_X(x) = \mathbf{P}(X=x)$ $p_{x,y}(x,y) = p_Y(y)p_{X\|Y}(x\|y)$ $\mathbf{P}(X=x) = \mathbf{P}(X \le x) - \mathbf{P}(X \le x-1)$	$\mathbf{P}_{X\|A}(x) = \mathbf{P}(X=x\|A)$ $\mathbf{P}_X(x) = \sum_y p_Y(y)p_{X\|Y}(x\|y)$	$\sum_x p_{X\|A}(x) = 1$

PDF:

$\mathbf{P}(X \in B) = \int_B f_X(x)dx$ $\quad$ $\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)dx$ $\quad$ $\mathbf{P}(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_x(x)dx = 1$

$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$ $\quad$ $\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$ $\quad$ $\mathbf{Var}(X) = \mathbf{E}[(X-\mathbf{E}[X])^2] = \int_{-\infty}^{\infty}(x = \mathbf{E}[X])^2 f_X(x)dx$

$0 \le \mathbf{Var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

$Y = aX + b \implies$ $\quad$ $\mathbf{E}[Y] = a\mathbf{E}[X] + b$ $\quad$ $\mathbf{Var}(Y) = a^2\mathbf{Var}(X)$

$\mathbf{P}(X \in B|A) = \int_B f_{X|A}(x)dx$ $\quad$ $Pr(f_{X|A}(x)) = \begin{cases} \frac{f_X(x)}{\mathbf{P}(X \in A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$ $\quad$ $\mathbf{P}(X \in B|X \in A) = \int_B f_{X|A}(x)dx$

$\mathbf{P}(X \in B|X \in A) = \int_B f_{X|A}(x)dx$ $\quad$ $\mathbf{E}[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$ $\quad$ $\mathbf{E}[g(X)|A] = \int_{-\infty}^{\infty} g(x)f_{X|A}(x)dx$

$disjoint\{A_1 ... A_n\} \implies f_x(x) = \sum_{i=1}^n \mathbf{P}(A_i)f_{X|A_i}(x)$ $\quad$ $\mathbf{E}[X] = \sum_{i=1}^n \mathbf{P}(A_i)\mathbf{E}[X|A_i]$ $\quad$ $\mathbf{E}[g(X)] = \sum_{i=1}^n \mathbf{P}(A_i)\mathbf{E}[g(X)|A_i]$

CDF:

$F_X(x) = \mathbf{P}(X \le x), \forall x$ $\quad$ $x \le y \implies F_X(x) \le F_X(y)$ $\quad$ $F_X(k) = \sum_{i=-\infty}^k p_x(i)$

$p_X(k) = \mathbf{P}(X \le k) - \mathbf{P}(X \le k-1) = F_X(k) - F_x(k-1)$

$F_X(x) = \int_{-\infty}^x f_x(t)dt$ $\quad$ $f_X(x) = \frac{dF_X}{dx}(x)$

$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ $\quad$ $f_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$

Uniform $f_X(x) = \frac{1}{b-a}$ $F_X(x) = \frac{x-a}{b-a}$

Expected Value and variance

$\mathbf{E}[X] = \sum_x x \cdot p_X(x)$	$Y = g(x) \implies \mathbf{E}[Y] = \sum_x g(x) \cdot \mathbf{P}_X(x)$
$\sigma^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2$	$Y = aX + b \implies \mathbf{E}[Y] = a \cdot \mathbf{E}[X] + b$ (linearity of expectation)
	$Y = X1 + ... + X_n \implies \mathbf{E}[Y] = \mathbf{E}[X_1] + ... + \mathbf{E}[X_n]$
	$\sum_{i=1}^N \mathbf{E}[X_i] = \sum_N \left(\sum_{i=1}^n \mathbf{E}[X_i]\right)\mathbf{P}(N=n)$

Conditioning/Unconditioning - $0 \le x_1 \le C, \quad 0 \le x_2 \le C$

$$\begin{aligned}\mathbf{E}[\ |X_1 - X_2|\ |X_2 = x_2] &= \mathbf{P}(X_1 > x_2)\mathbf{E}[X_1 - x_2|X_1 > x_2] + \mathbf{P}(X_1 < x_2)\mathbf{E}[x_2 - X_1|X_1 < x_2] \\ &= \left(\frac{C - x_2}{C}\right)\left(\frac{C - x_2}{2}\right) + \left(\frac{x_2}{C}\right)\left(\frac{x_2}{2}\right)\end{aligned}$$

$$\mathbf{E}[|X_1 - X_2|] = \int_0^C \mathbf{E}[|X_1 - X_2||X_2 = x2]f_{X_2}(x_2)dx_2$$

Markov:	$\mathbf{P}(X \ge a) \le \frac{\mathbf{E}[X]}{a}$	**Chebyshev:**	$\mathbf{P}(\|X - u\| \ge c) \le \frac{\sigma^2}{c^2} \quad \forall c > 0$

Markov Chain

$$p_{ij} = \mathbf{P}(X_{n+1} = j|X_n = i) \qquad \sum_{j=1}^m p_{ij} = 1$$

$$\pi_j = \sum_{k=1}^m \pi_k p_{kj} \qquad \sum_{k=1}^m \pi_k = 1$$

$\pi_j = 0 \implies$ transient state $\quad$ $\pi_j \ge 0 \implies$ recurrent state

$\mathbf{P}(X_0 = j) = \pi_j(X_0$ is initial state then):

$$\mathbf{P}(X_1 = j) = \sum_{k=1}^m \mathbf{P}(X_0 = k)p_{kj} = \sum_{k=1}^m \pi_k p_{kj} = \pi_j$$

Poisson process:

Small interval probabilities

$\mathbf{P}(0,\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau)$

$\mathbf{P}(1,\tau) = \lambda\tau e^{-\lambda\tau} = \lambda\tau(1-\lambda\tau) = \lambda\tau - \lambda^2\tau^2 = \lambda\tau + o_1(\tau)$

$\mathbf{P}(k,\tau) = 0 + o_k(\tau)$

Large interval probabilities

$$p_{N_\tau}(k) = \mathbf{P}(k,\tau) = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}, \quad k = 0, 1, ...$$

$\mathbf{E}[N_\tau] = \lambda\tau$

Time T of the first arrival

$F_T(t) = \mathbf{P}(T \le t) = 1 - \mathbf{P}(T > t) = 1 - \mathbf{P}(0,t) = 1 - e^{-\lambda t}$

$f_T(t) = \lambda e^{-\lambda t}, \quad t \ge 0$