Conditional)

$P(B|A) \to P(A\cap B)$; $P(\bar{B}|A) \to P(A\cap\bar{B})$; $P(B|\bar{A}) \to P(\bar{A}\cap B)$; $P(\bar{B}|\bar{A}) \to P(\bar{A}\cap\bar{B})$

$\mathbb{P}[A|B] = \mathbb{P}[A\cap B]/\mathbb{P}[B]$
$\quad = \mathbb{P}[B|A]\,\mathbb{P}[A]/\mathbb{P}[B]$

$\mathbb{P}[A\cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A|B]\,\mathbb{P}[B]$

$\mathbb{P}[A\cap B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A\cup B] = \mathbb{P}[A|B]\,\mathbb{P}[B]$

NAME	Normal(μ,σ^2)	Exponential(λ)	Uniform(a,b)	Poisson(λ)	Geometric(p)	Binomial(n,p)	Bernoulli(p)	NAME
	$\mu\in\mathbb{R},\ \sigma^2>0$	$0<\lambda<\infty$	$a\le b,\ n:=b-a+1$	$0<\lambda<\infty$	$p\in[0,1]$	$n\in\mathbb{Z}^+,\ p\in[0,1]$	$p\in[0,1]$	
Ex		Time between poi. events	Equal p. $\forall\omega$	# events in period; $\sim Bin(n,p)$ | $\sim Ber(n/p)$	# trials til succ	# succ. of n indep trials	1 flip, n=1 Bin.	Ex.
$\mathbb{E}[X]$	μ	$1/\lambda$	$\frac{a+b}{2}$	λ	$1/p$	np	p	$\mathbb{E}[X]$
$Var(X)$	σ^2	$1/\lambda^2$	$\frac{n^2-1}{12}$	λ	q/p^2	$np(1-p)$	$p(1-p)=pq$	$Var(X)$
PMF $=p_X$	$\frac{\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}{\sqrt{2\pi\sigma^2}}$	$\lambda e^{-\lambda x}$	$\frac{1}{b-a+1}=\frac{1}{n}$	$\frac{\lambda^k}{k!}e^{-\lambda}$	$p(1-p)^{k-1}$	$\binom{n}{k}p^k(1-p)^{n-k}$	$\begin{cases} q: & k=0 \\ p: & k=1\end{cases}$	PMF $=p_X(x)$
MGF	$\exp(\mu t + \sigma^2t^2/2)$	$\frac{\lambda}{\lambda-t}$	$\frac{e^{at}-e^{bt}}{n(1-e^t)}$	$\exp(\lambda(e^t-1))$	$\frac{pe^t}{1-qe^t}$	$(q+pe^t)^n$	$q+e^t\cdot p$	MGF
	$\frac{X-\mu}{\sqrt{\sigma^2}}\sim N(0,1)$	CDF $=1-e^{-\lambda x}$; Memoryless: $\mathbb{E}[X\mid X>3]=3+\mathbb{E}[X]$	CDF: $\frac{\lfloor x\rfloor-a-1}{b-a+1}$	$X,Y\sim Poiss(\lambda,\mu)$, $\underbrace{X+Y}_{\text{indep}}\sim P(\lambda+\mu)$	CDF: $1-q^{\lfloor x\rfloor}$	$X,Y\sim Bin(n/m,p)$, $\underbrace{X+Y}_{\text{indep.}}\sim Bin(n+m,p)$		Misc.

Axioms:

- Ω = sample space, set of samples
- $\mathcal{F}$ = Events, subsets of Ω
  - $\hookrightarrow$ σ-alg:
    1) Non-empty
    2) Closed under compliment
    3) " " countable unions
- $\mathbb{P}$ = Probability measure
  - $\hookrightarrow \mathcal{F}\to[0,1]$, Kolmogorov:
    1) Non-neg: $\forall\omega\in\mathcal{F}: \mathbb{P}(\omega)\ge 0$
    2) $\mathbb{P}[\Omega]=1$
    3) Countable Additivity: for disjoint

$1 = \mathbb{P}[[0,1]] = \sum_{x\in[0,1]}\mathbb{P}[\{x\}] = 0$

$\mathbb{P}[\bigcup A_i] = 1 - \mathbb{P}[\bigcap A_i^c]$

Union b.: $\le \Sigma^n\,\mathbb{P}[A_i]$ (ignore dependence)

Independence:

$\mathbb{P}[A\cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$

$\mathbb{P}[A|B] = \mathbb{P}[A]$

$\mathbb{E}[AB] = \mathbb{E}[A]\,\mathbb{E}[B]$

$Cov(A,B) = \mathbb{E}[AB] - \text{''} = 0$

$Var(A+B) = Var(A) + Var(B)$

$M_{A,B}(s,t) = M_A(s)M_B(t)$

Conditionally indep: A,B given C

$\mathbb{P}[A\cap B|C] = \mathbb{P}[A|C]\,\mathbb{P}[B|C]$

$\mathbb{P}[A\cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A\cap B]$

Inclusion Exclusion $= \mathbb{P}[A\setminus\{A\cap B\}] + \mathbb{P}[B]$

$\mathbb{P}[\bigcup A_i] = \sum_{J\subseteq\{1\ldots n\}}(-1)^{|J|+1}\bigcap_{j\in J}A_j$

Derange

$= \sum_{k=1}^{n}(-1)^{k+1}\binom{n}{k}\frac{(n-k)!}{n!}$

$= \sum(-1)^{k+1}\cdot 1/k!$

Let $p_i := \mathbb{P}[B_i]$ be none in assigned set:

$p_0=1,\ p_1=0,\ n\ge 2\ldots$

$p_n = \frac{n-1}{n}p_{n-1} + \frac{1}{n}p_{n-2}$

$d_n := p_n - p_{n-1}$

$= \ldots p_{n-1} = -\frac{1}{n}d_{n-1}$

$\therefore p_n = (p_n - p_{n-1}) + \ldots (+p_0)$

$= \sum_{i=0}^n d_i$

$= \sum(-1)^i/i!\ \square$

Conc. Ineq.

$\mathbb{E}[X] \ge \mathbb{E}[\lambda\cdot\mathbb{1}\{X\ge\lambda\}]$
$\quad = \lambda\,\mathbb{P}[X\ge\lambda]$

$\therefore \mathbb{P}[X\ge\lambda] \le \frac{\mathbb{E}X}{\lambda}$ } Markov [non-neg]

If X sym: $\mathbb{P}[X\ge\mathbb{E}[X]+\varepsilon] = $

$\overset{t>0}{=} \mathbb{P}[\overbrace{e^{X t}\ge e^{\lambda t}}^{X\ge\lambda}]$

$\le \min_t \frac{M_X(t)}{e^{t\cdot\lambda}}$ } Chernoff

$2\mathbb{P}[X\ge\mathbb{E}[X]+\varepsilon]$
$+\mathbb{P}[X\le\mathbb{E}[X]-\varepsilon]$

$\mathbb{P}[|X-\mathbb{E}X|\ge\varepsilon]$
$\le \frac{Var(X)}{\varepsilon^2}$ } Chebyshev

Total Prob:

$\mathbb{P}[B] = \mathbb{P}[B\cap\overbrace{\Omega}^{\bigcup A_i}]$

$= \mathbb{P}[\bigcup(B\cap A_i)]$

$= \sum_{i=1}^n\mathbb{P}[B\cap A_i]$

$= \sum_{a\in A}p_{A,B}(a,b)$

LOTUS:

$Y = g(X)$
$g: X\to\mathbb{R}$

$\mathbb{E}[Y] = \sum_y y\,p_Y(y)$

$= \sum_y y\sum_{x: g(x)=y}p_X(x)$

$= \sum_y\sum_x g(x)p_X(x)$

$= \sum_{x\in X}g(x)p_X(x)$

BDay:

$\mathbb{P}[\bigcap A_i]$

$\overset{\text{Indep}}{=} \prod_{i=1}^{n}(1-\frac{i-1}{365})$

$1-x \le e^{-x}$: $\le \prod\exp(-\frac{i-1}{365})$

$= \exp(\Sigma\ \text{''})$

$= \exp\left(\binom{n}{2}/-365\right)$

Tail-sum

$\mathbb{E}[X] = \sum_{k\ge 1}k\cdot p_X(k)$

$\sum_{k\ge 1}\sum_{j=1}^{k} = \sum_{j\ge 1}\sum_{k\ge j}$

$= \sum_{j\ge 1}\mathbb{P}[X\ge j] = \int_0^\infty p_X(x)\,dx$

$Var(X) = \mathbb{E}[(X-\mathbb{E}[X])^2]$
$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$
$= \sum_x(x-\mathbb{E}[X])^2p_X(x)$

Tower:

$\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)\underbrace{\mathbb{E}[X|Y]}]$

Func of $y = g(y) \sim \sum_x x\,p_{X|Y}(x|y)$

$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$

$f(Y) = 1$ | Iterated Exp.

$Cov(X,Y) =$
$\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$
$= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$
$= 0 \iff$ uncorrelated

indep $\Longrightarrow$ " ; " $\not\Leftarrow$ " | $X\sim U[-1,1]$, $Y=X^2$

Coupon:

X_i := # boxes until i^{th} coupon (having $i-1$)

$\sim Geo(N-i-1/N)$

$\mathbb{E}[\#\text{boxes}] = \sum_{i=1}^N\mathbb{E}[X_i] = \frac{N}{N} + \frac{N}{N-1} + \ldots + \frac{N}{1} = NH_N$

$\approx N\log(N)$

Continuous R.V.s

$\mathbb{P}[X\in B] = \int_B\underbrace{f_X(x)}\,dx$

$\ldots\ge 0\ \forall x$

$\int_{\mathbb{R}}\ldots = 1$

$F_X(x) = \mathbb{P}[X\le x]$
$= \int_{-\infty}^x f_X(u)\,du$

$f_X(x) = \int_{-\infty}^{\infty}f_{XY}\,dy$

$f_{X|Y}(x|y) = f_{XY}(x,y)/f_Y(y)$

$\mathbb{E}[X|Y] = \int x\,f_{X|Y}(x|y)\,dx$

Memoryless: $\mathbb{P}[X>s+t\mid X>s] = \mathbb{P}[X>t]$

$\therefore\mathbb{P}[X>s+t] = \mathbb{P}[X>s]\,\mathbb{P}[X>t]$

$\Longrightarrow F_X(t) = 1-\mathbb{P}[X>t] = 1-e^{\lambda t},\ |\lambda|>0$

$Var(X|Y) = \mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2$

$\mathbb{E}[\ \text{''}\ ] = \mathbb{E}[(X-\mathbb{E}[X|Y])^2]$ } MMSE

$Var(X) = \mathbb{E}[Var(X|Y)] + \underbrace{Var(\mathbb{E}[X|Y])}$ } Total Variance

How much uncertainty of X is reduced by knowing Y, on avg.

Derived Dist.

Given $Y=g(X)$, we can either use LOTUS and op. on dist of Y or CDF: $\overbrace{g(X)\le y}$

$F_Y = \mathbb{P}[Y\le y]$

$= \mathbb{P}[X\in g^{-1}(-\infty,y]]$

$\{x \mid g(x)\le y\}$

$p_Y(y) = \sum_{x: g(x)=y}p_X(x)$

Ex: $f_X(x) = \{1: x\in[0,1)$
$Y = 2X$

$f_Y = \frac{d}{dy}\underbrace{F_Y}_{F_X(y/2)} = \frac{1}{2}f_X(\frac{y}{2})$

$= \{\frac{1}{2}: y\in[0,2)$

$F_{X^{(i)}}(x) = \sum_{k=i}^n\binom{n}{k}F(x)^k(1-F(x))^{n-k}$

Order:

$f_{X^{(i)}}(y)\,dy \approx \mathbb{P}[X^{(i)}\in\overbrace{[y,y+dy]}^{\gamma}]$

$\therefore f_{X^{(i)}}(y) = \underbrace{n}_{X_i\text{ fall in }\gamma}\underbrace{\binom{n-1}{i-1}}_{\text{choose } i-1 \text{ of } n-1 \text{ left} \le y}\underbrace{F_X(y)^{i-1}}_{\text{prob}}\underbrace{(1-F_X(y))^{n-i}}_{\text{prob not}}\underbrace{f_X(y)\,dy}_{\text{prob one } X_i\in\gamma}$

MGF | Uniquely describes dist.

$M_X(t) := \mathbb{E}[e^{tX}]$

$= \mathbb{E}\left[\sum_{n\ge 0} \frac{(tX)^n}{n!}\right]$

$= \sum_{n\ge 0} \frac{t^n}{n!} \underbrace{\mathbb{E}[X^n]}_{n^{th}\text{ moment}}$

$M_X(0) = \mathbb{E}[e^{0}] = 1$

$\mathbb{E}[X^p]$

$= \left[\frac{d^p}{dt^p} M_X(t)\right]_{t=0}$

$= \left[\sum_{k\ge p}^{\infty} \frac{t^{k-p}}{p!} \mathbb{E}[X^k]\right]_{t=0}$

Sampling: Want to simulate random sampling from dist. X.

$U \sim U[0,1]$, $X \sim E(\lambda)$ st.:

$u_0 = F_X^{-1}(U)$ is our random sample.

Pf: $Y = F_X^{-1}$: $CDF_Y = \mathbb{P}[Y \le y]$

$= \mathbb{P}[U \le F_X(y)]$

$= F_X(y)$ □

Polling: $n = 10^6$, $\bar{\mu} = \frac{1}{2}$, $\bar{\sigma} = \frac{3}{16}$

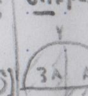
WTF: $O(\#$ w/ score $\ge 1)$

$\mathbb{P}[\bar{\mu} \ge 1] = \mathbb{P}[(\bar{\mu} - \frac{1}{2}) \ge (1 - \frac{1}{2})]$

$\le \frac{Var(\bar{\mu})}{1/4}$ $\quad \mathbb{E}[\bar{\mu}]$, $\bar{\mu} = \frac{1}{2}$, $\frac{1}{4}$

$\le \frac{3}{4} \therefore 10^6 \cdot \frac{3}{4}$ is v. bound

Graphical $F_X = \int_{\sqrt{1-x^2}}^{0} 3A\,dy + \int A\,dy$



$x \ge 0$

$x = \frac{2}{\pi}\sqrt{1-x^2}$

Same for $x < 0$ via sym.

Same if $x \overset{d}{=} Y$ via rot sym.

Since F_X is even, $F(x) = F(-x)$

Indep $\Rightarrow F(x)F(y) = F(x,y)$

$F(-x)F(y) =$

$\ne \Rightarrow$ dependent

$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ — 0 via reflection

Let $I := $ 1st quad.

$\int_0^1 \int_0^{\sqrt{1-x^2}} xy\, F_{XY}(x,y)\,dy\,dx$

st. $\mathbb{E}[XY] = I - 3I + I$

$- 3I = -4I$

$-4\int_0^1 A\left(x\left[\frac{y^2}{2}\right]_0^{\sqrt{1-x^2}}\right) = -\frac{1}{4\pi} \ne 0$

$\therefore$ correlated

Non-seq Events:

Each round we have choices $\{A,B,C,D\}$, not random and no repeats.

$\mathbb{P}[A_1] = 1/4$

$\mathbb{P}[A_i] = 1/3(1 - \mathbb{P}[A_{i-1}])$

$\sum_{i=1}^{n} a_1 \cdot r^{i-1} = \frac{a_1(1-r^n)}{1-r}$

$S^T S_{ij} = \sum_{k=1}^{d} S_{k,i} S_{k,j}$, $S \in \mathbb{R}^{d\times n}$

$i = j$ $\mathbb{E}[\ ]_{ij} = \mathbb{E}[\ ] = d\,\mathbb{E}[s_{ki}^2]$

$Var(\ ) = Var(\ ) = d\,Var(\ )$

$i \ne j$ $\mathbb{E}[\ ] = \mathbb{E}[s_{ki}s_{kj}] = \mathbb{E}[s_{ki}]\mathbb{E}[s_{kj}]$

$Var(\ ) = d\,Var(\ ) = d\,\mathbb{E}[s_{ki}^2]^2$ indep

Exponentials

$X \sim E(\lambda)$, $Y \sim E(\mu)$, indep.

$\mathbb{P}[X < Y] = \int_0^\infty \underbrace{\mathbb{P}[X \le Y \mid Y = y]}_{1-\exp(-\lambda y)} \underbrace{F_Y(y)}_{\mu\exp(-\mu y)}\,dy = \frac{\lambda}{\lambda+\mu}$

$X = \min X_i$

$\mathbb{P}[X > x] = \prod_{n=1}^{n} \underbrace{\mathbb{P}[X_n \ge x]}_{\exp(-\lambda_n x)} = \exp(-x\sum_{n=1}^{n}\lambda_i)$ $\therefore X = \min X_i \sim Exp(\sum \lambda_i)$

$\mathbb{P}[X_i = \min_k X_k] = \mathbb{P}[X_i \le \min_{k\ne i} X_k] = \lambda_i / \sum_{j=1}^{n} \lambda_j$ □

$Y = \max X_i$

$\mathbb{P}[Y \ge y] = F_Y(y) = \prod_{i=1}^{n} \mathbb{P}[X_i \le y] = (1 - e^{-\lambda y})^n \therefore f_Y = \lambda n e^{-\lambda y}(1-e^{-\lambda y})^{n-1}$

Median: Given $2n+1$ samples. WTF $\mathbb{E}[X_{(n)}]$

$X_{(k+1)} - X_{(k)}$ is the interval w/ $2n-k$ left, memoryless

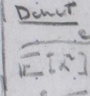
$\sim Exp(\lambda(2n-k))$

$\therefore \mathbb{E}[X_{(n)}] = \sum_{k=0}^{n-1} \underbrace{\mathbb{E}[X_{(k+1)} - X_{(k)}]}_{1/\lambda(2n+1-k)} = \frac{1}{\lambda}\sum_{k=n+1}^{2n+1} 1/k$

$H_n = \frac{1}{1} + \dots \frac{1}{n}$ $\quad = \frac{1}{\lambda}(H_{2n+1} - H_n)$

exp. of median

Donut $X : x$ on eq. sym



$\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2] = \mathbb{E}[R^2]$

$\therefore \mathbb{E}[X^2] = \frac{1}{2}\mathbb{E}[R^2]$

$Var(X)$, as $\mathbb{E}[X] = 0$ via sym

$\mathbb{P}[R \le r] = \frac{\pi(r^2 - 1/4)}{\pi(1 - 1/4)} = \frac{4}{3}r^2 - \frac{1}{3}$

$f_R(r) = \frac{d}{dr}(\ ) = \frac{8}{3}r$

$\therefore \mathbb{E}[R^2] = \int_{1/2}^{1} r^2 f_R(r)\,dr$

$= \frac{2r^4}{3}\Big|_{1/2}^{1} = \frac{5}{8}$

$\therefore Var(X) = \frac{5}{16}$

Convergence |

Almost-sure: $X_n \xrightarrow{a.s.} X$ if

$\mathbb{P}[\lim_{n\to\infty} X_n = X] = 1$

$\{\omega : \lim X_n(\omega) = X(\omega)\}$

In-prob: $\xrightarrow{P}$ $\forall \epsilon > 0$

$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0$

In-dist: $\xrightarrow{d}$ $\forall x : p_X(x) = 0$ continuity pts

$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$

Min Geo. $\mathbb{P}[\underbrace{X > K, Y > K}_{\min(X,Y) > K}] = \underbrace{\mathbb{P}[X > K]\mathbb{P}[Y \ge K]}_{indep.} = \underbrace{(1-p)^K(1-q)^K}_{\sim Geo(1-(1-p)(1-q))}$

Min Extra: $\mathbb{E}[\min(X,3)] = 3\,\mathbb{P}[X > 3] + \mathbb{P}[X \le 3]\mathbb{E}[X \mid X \le 3]$

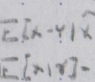
$X := \min(E(\lambda_1), E(\lambda_2))$ $\quad = \int_0^\infty \mathbb{P}[X_1 > t]\,\mathbb{P}[X_2 > t]\,\mathbb{P}[3 > t]\,dt$

$= \frac{1 - e^{-3(\lambda_1+\lambda_2)}}{\lambda_1+\lambda_2}$ □ $\quad = \int_0^3 e^{-\lambda_1 t} \cdot e^{-\lambda_2 t}\,dt = \frac{-1}{\lambda_1+\lambda_2}\left[e^{-t(\lambda_1+\lambda_2)}\right]_0^3$

$\mathbb{E}[X - Y \mid X + Y]$

$= \mathbb{E}[X \mid \sigma] - \mathbb{E}[Y \mid \sigma]$

$= \frac{X+Y}{2} - \frac{X+Y}{2} = 0$



$X + Y$ parallel to $1 - x$

$\Rightarrow X - Y \perp$ and uniform dist

Apples

$R \sim N(a, 2)$, $M \sim N(R^2, 4)$,

N: # picked $\sim G(p)$: T = total mass $= \sum_{i=1}^{N} M_i$

$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T \mid N]]$

$= \mathbb{E}[\mathbb{E}[\sum_{i=1}^{N} M_i]]$

$= \mathbb{E}[N \cdot \mathbb{E}[M_i]]$

$\mathbb{E}[\mathbb{E}[M_i \mid R]] = \mathbb{E}[R^2] : Var(M) + \mathbb{E}[R]^2$

$\mathbb{E}[T] = \mathbb{E}[N(2 + a^2)] = 2 + a^2$

$= (2+a^2)\mathbb{E}[N] = (2+a^2)/p$ □

$(X_n \xrightarrow{a.s.} X) \Rightarrow (X_n \xrightarrow{P} X) \Rightarrow (X_n \xrightarrow{d} X)$

LLN | $\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mathbb{E}[X]$ — WL Doesn't give details on any one X_i's behav

$\mathbb{E}[X] < \infty$ — in p, says weak

$X_i \sim X$ iid — a.s., says strong

Strong law says $\forall \omega : \mathbb{P}(\omega) = 1$, $M_n(\omega) \to \mathbb{E}[X]$

In-L^p: $\lim_{n\to\infty} \mathbb{E}[|X_n - X|] = 0$

WTS: $X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{P} X$

$\mathbb{P}[|X_n - X| \ge \epsilon] \overset{m}{\le} \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p}$

$= 0$ □

Ant: $B_0 = 0$, $B_i = 1 + pB_{i+1} + qB_0$

$a = \frac{1}{p}$, $b = -\frac{1}{p} - \frac{(1-p)B_0}{p}$

$\Rightarrow B_{i+1} = aB_i + b$

s.t. $B_i = a^i B_0 + \frac{1-a^i}{1-a} b$

$B_n = 0$: $i = n$

$\therefore B_0 = \frac{1-p^n}{p^n - p^{n+1}}$

ALT: N : # failed revisit 0 $\sim Geo(p^n) = \mathbb{P}_0[T_n < T_0^+]$

$T_k \overset{iid}{=} K^{th}$ return time to 0

$\mathbb{E}[T_1] = \mathbb{E}_0[T_0^+] = 1/\pi(0)$

$\pi_0 = q\sum_{i=0}^{n} + p\pi_n$ $\left\{ = \frac{1-p^{n+1}}{1-p}\right.$

$= q + p^{n+1}\pi_0$

$\mathbb{E}[\sum^{N} \tau_{n-1}] = \mathbb{E}[N]\mathbb{E}[\tau_1] - 1$

$= \frac{1}{p^n} \cdot \frac{1-p^{n+1}}{1-p}$

BBall: Rates i.i.d λ_c, λ_s : End @ K pts ahead

Q1: Let $\lambda_c = \lambda_s$ and c starts $m < K$

$\mathbb{P}[C \text{ wins}] = ?$ | Let $p = \lambda_c / (\lambda_c + \lambda_s) = \frac{1}{2}$

$P_{c,i} = 1$, $P_{n,n-1} = P_{n,n+1}$, $P_{i,i+1} = p$, $P_{i,i-1} = 1-p$

$\alpha_0 = 0$, $\alpha_{2K} = 1$, $\alpha_i = \frac{1}{2}(\alpha_{i-1} + \alpha_{i+1})$

$\therefore \alpha_{m+K} = \frac{m+K}{2K}$ — Evenly distributed sym. Gambler's : $\alpha_i = \frac{i}{n}$

Q2: $\mathbb{E}[T] = ?$

$= \mathbb{E}[\sum_{i=1}^{N_j} \tau_n] = \mathbb{E}[N_j \cdot \mathbb{E}[\tau_1]]$

$= \mathbb{E}[N_j]\,\mathbb{E}[\tau_1]$; [# of visits] $\quad \frac{1}{2(\lambda_c + \lambda_s)}$, $j = m+K$

$B_i := \mathbb{E}[N_i]$; $B_0 = B_{2K} = 0$

$\Delta_i := B_{i+1} - B_i$

$\Rightarrow \Delta_i = \Delta_{i-1} - 2$ $\quad -B_{2K-1} = \Delta_{2K-1}$

$\Delta_{2K-1} = \Delta_0 - 2(2K-1)$ $\quad = \Delta_0 - 2(2K-1)$

$B_{2K-1} = B_1 = 2K - 1 + \Delta_0$ $\quad = B_1 - 2(2K-1)$

$\Rightarrow \Delta_i = \Delta_0 - 2i = 2K - 1 - 2i$

$\therefore B_j = B_0 + \sum_{i=0}^{j-1} \underbrace{B_{i+1} - B_i}_{\Delta_i = i} = j(2K - j)$

$\therefore B_{m+K} = (K+m)(K-m)$

Inversion: If $x_i < x_j$ and $i < j$

A_{ij} := event i appears before j, $|A_{ij}| = \binom{n}{2}$

$\mathbb{E}[X] = \sum \mathbb{E}[A_{ij}] = \binom{n}{2}\mathbb{P}[A_{ij}] = \frac{1}{2}\binom{n}{2}$ # pair relations

CLT | $X_i \overset{iid}{\sim} X \sim N(\mu, \sigma^2)$

$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} \xrightarrow{d} N(0,1)$

$\mathbb{P}[|\hat{p} - p| \ge \epsilon] \le \delta$

$S_n/n = \hat{p}$ $\therefore$ $\quad \overset{c}{\le} Var(\frac{S_n}{n} - p)/\epsilon^2$

$n = \frac{p(1-p)}{\delta\epsilon^2} \Leftarrow \left\{ = p(1-p)/n\epsilon^2 \right.$ $= \delta$

$\epsilon \to \epsilon/2 \Rightarrow n \to 4n$

$\delta \to \delta/2 \Rightarrow n \to 2n$

Stick Break $n \ge 1$ times, keep $X_i \overset{iid}{\sim} U[0,1]$

$R = \prod_{i=1}^{n} X_i$ be total we have @ n

$\lim_{n\to\infty} \ln P_n^{1/n}$

$= \lim \frac{1}{n}\sum \ln(X_i)$

$\overset{a.s.}{=} \mathbb{E}[\ln(X_i)]$

$= \int_0^1 \ln x\,dx = -1$

$\therefore \mathbb{P}[\ln P_n^{1/n} \to -1]$

$= \mathbb{P}[P_n^{1/n} \to e^{-1}] = 1$

Confident Gambling

$\hat{p} = \sum_{i=1}^{N} X_i / N$

$X_i = \mathbb{1}\{\text{roll } 6\}$

Q: 95% conf. int. for p

$\hat{p} \pm \frac{1}{2\sqrt{N}}\Phi^{-1}(0.975)$

Q: Suppose $\hat{p}$ w/in true by 0.01 w.p. .95

Find N:

$\frac{1}{\sqrt{N}} \le 0.01 \Rightarrow N \ge 10{,}000$

Coupon Bound $\mathbb{E}[X] = n \cdot H_n$

$X_i \sim G(\frac{n-i+1}{n})$ $\quad H_n \in [\ln(n), \ln(n+1)]$

$\mathbb{P}[X > 2nH_n] \overset{M}{\le} \mathbb{E}[X]/2nH_n = 1/2$

$\mathbb{P}[\ ] \le \mathbb{P}[|X - nH_n| > nH_n] \overset{c}{\le} \frac{Var(X)}{(nH_n)^2} \le \frac{\sum(\frac{n^2}{i^2})}{}$ $\le \frac{\pi^2}{6}$

A_i = Fail to get i after $\mathbb{E}[X]$ $\quad = \frac{\pi^2}{6\ln(n)^2}$

$\mathbb{P}[A_i] \le \left(\frac{n-1}{n}\right)^{2nH_n} = \left(\left(1 - \frac{1}{n}\right)^n\right)^{2H_n} \le e^{-2H_n} \le e^{-2\ln(n)} = 1/n^2$

MARKOV CHAINS

Row π_i represents [$i\to1$ $i\to2$ ... $i\to N$]

Rows sum to $1 = \sum \pi_i$ $\forall i\in S$

Chapman-Kolmogorov:

- $\mathbb{P}[X_n=j \mid X_0=i] = p_{ij}^{(n)} = [P^n]_{ij}$
- $\mathbb{P}[X_0=x_0, X_1=x_1, \dots, X_n=x_n] = \pi_0(x_0)P(x_0,x_1)\cdots P(x_{n-1},x_n)$

Stationary Dist: $\pi = (\pi_i)_{i\in S} = [\pi_1 \cdots \pi_n]$

$\pi_j = \sum_{i\in S}\pi_i P_{ij} = 1/\mathbb{E}[T_j \mid X_0=j]$

- Balance Eq: $\pi = \pi P$

↪ All but 1 eqn linearly indep; use $\sum \pi_i = 1$

Classification of States

- $i\to j \iff \exists$ path from $i\to j$
- $i\to j,\ j\to i \iff i\leftrightarrow j$; i,j communicate

↪ Any property for i holds for j

↪ $i\leftrightarrow k, k\leftrightarrow j \iff i\leftrightarrow j$

Recurrence: $T_i = \min_{n\ge1}\{X_n = i\}$

# of steps until returning

Properties:

- Irreducible: One class, S

↪ Ex. connected graph; all can communicate

- Recurrent: Given $X_0=i$, we re-visit i w.p. 1

↪ revisit once ⇒ we revisit i ∞ times w.p. 1

◦ Positive: $\mathbb{E}_i[T_i \mid X_0=i] < \infty$

↪ $\forall$ irreducible finite MC

◦ Null: $\mathbb{E}_i[T_i \mid X_0=i] = \infty$

- ¬ Recurrent $\iff$ Transient

↪ Any state w/ transition to another class

Ex| Birth-Death Chain

$p < 1/2$: positive

$p > 1/2$: transient

$p = 1/2$: null

BIG | Let $(X_n)_{n\ge0}$ be irreducible MC

1.) All states transient or null recurrent

$\iff \nexists\, \pi$

$\implies \lim_{n\to\infty} P_{ij}^n = 0 \quad \forall i,j\in S$

Irreducible finite MC

2.) All states Positive Recurrent

$\iff \exists$ unique π

$\implies \pi_j = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} P_{ij}^k = 1/\mathbb{E}[T_j \mid X_0=j]$

~ if aperiodic $\implies \lim_{n\to\infty} P_{ij}^n = \pi_j$

CTMCs: $\pi_j = \lim_{t\to\infty} P_{ij}^t = 1/q_j \cdot \mathbb{E}[T_j \mid X_0=j]$

Period: $\gcd_{n\ge1}\{P_{ii}^n > 0\}$

↪ If starting at i, we only possibly revisit at integer multiples of period

↪ Class property

↪ Aperiodic: period$(i) = 1$; self-loop

CTMC | Characterized by Q:

1.) $[Q]_{ij} \ge 0 \quad \forall i\ne j \in S$

2.) $\sum_j [Q]_{ij} = 0 \quad \forall i,j\in S$

↪ $[Q]_{ii} = -\sum_{j\ne i}[Q]_{ij} = -q_i = -\sum_{j\ne i}\lambda_{ij}$

w/ $[Q]_{ij} = q_i P_{ij}$

$\pi Q = 0$

w/ $\sum \pi_i = 1$

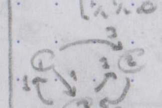
$\pi_i q_i = \sum_{j\ne i}\pi_j Q_{ji} = \sum \pi_j q_j P_{ji}$

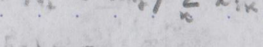
- Hold i for $\mathrm{Exp}(q_i) = \mathrm{Exp}(\sum_j \lambda_{ij}) = \min_j\{\mathrm{Exp}(\lambda_{ij})\}$ $= \frac{1}{\sum_j \lambda_{ij}}$

- Jump $i\to j$ w.p. $P_{ij} = \lambda_{ij}/\sum_k \lambda_{ik}$

Ex| $Q = $ [illegible]

$\therefore P = $ [illegible]

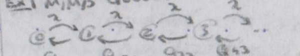


Ex| 

$Q_{i,i+1} = \lambda \quad \forall n\ge0$

$Q_{i,j} = 0$ o.w.

Ex| M/M/S Queue:



$q_{i,i+1} = \lambda$

$q_{i,i-1} = \begin{cases} i\mu & i\in[1,s] \\ s\mu & i > s \end{cases}$

Customers arrive $\sim$ PP(λ)

Served if any of S available

Service takes $\sim$ Exp(μ)

Ex| Birth $\sim$ PP(λ)

Live $\sim$ Exp(μ)

$Q_{i,i+1} = i\lambda$, $Q_{i,i-1} = n\mu$ | $P_{n,n+1} = \lambda/\lambda+\mu$

Reversible: $\pi_i P_{ij} = \pi_j P_{ji} \Rightarrow \pi_j \sum_i P_{ji} = \sum_i \pi_i P_{ij}$

- Suffices if [transition diagram is made undirected & self-loops rm'd] is a Tree

- If $X_i \sim \pi$: $(X_0 \cdots X_n) \overset{d}{=} (X_n \cdots X_0)$

$P_{ji} = \frac{\pi_i}{\pi_j} P_{ij}$

$\overbrace{\sum_{i\ne j} \pi_i q_{ij} = \pi_j(-q_j)}^{\pi Q = 0}$

Metropolis Hastings 1.) Propose next $y \sim g(x,\cdot)$

2.) Accept w.p. $A(x,y)$

$A(x,y) = \min\left\{1, \frac{f(y)\,g(y,x)}{f(x)\,g(x,y)}\right\}$; $P(x,y) = g(x,y)A(x,y)$

Q| WTS $p(x)$ holds DBE for $P(x,y)$

wlog: $f(y)g(y,x) \le f(x)q(x,y)$ s.t. $1 = A(y,x)$

$p(x)P(x,y) = p(x)g(x,y)\cdot \underbrace{\frac{f(y)g(y,x)}{f(x)g(x,y)}}_{A(x,y)}$

$= p(x)\frac{f(y)}{f(x)}g(y,x)$

$= p(y)\ g(y,x)\cdot \underbrace{1}_{P(y,x)}$ ∎

Jump Chain: Given by transition matrix P

$\underbrace{\pi}_{\text{CTMC}} \ne \underbrace{\tilde{\pi}}_{\text{Jump chain}} \rightsquigarrow \tilde{\pi}_i = \pi_i q_i / \sum_j \pi_j q_j$

$\pi_i = \tilde{\pi}_i/q_i \big/ \sum_j \tilde{\pi}_j/q_j$

Ex| M/M/∞ Queue

$q_{n,n+1} = \lambda$ arrivals $\sim$ PP(λ)

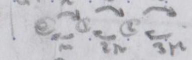
$q_{n,n-1} = n\mu$ service time min iid Exp(μ)

$\pi Q = 0 \Rightarrow \pi_n = e^{-r}\cdot r^n/n!$, $r = \lambda/\mu$ | X_t = # people in queue at t

i.e. $X_t \xrightarrow{d} \mathrm{Pois}(r)$

Ex| Arrive $\sim$ PP(λ)

Stay $\sim$ Exp(μ)



Q: π? Insight:

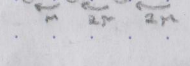
DBE: $\pi_n \overbrace{q_{n,n-1}}^{n\mu} = \pi_{n-1} \overbrace{q_{n-1,n}}^{\lambda}$

$\implies \pi_n = \frac{\lambda}{\mu n}\pi_{n-1} = \left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{n!}\underbrace{\pi_0}_{C}$

$1 = \sum_{n=0}^{\infty}\pi_n = \sum\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} C = e^{\lambda/\mu}\cdot C$

$\therefore C = \pi_0 = e^{-\lambda/\mu}$

Ex M/M/2 Queue



$\pi_1 = \frac{\lambda}{\mu}\pi_0$

$\pi_{i+1} = \frac{\lambda}{2\mu}\pi_i$

$\sum_{i=0}^{\infty}\pi_i = \pi_0 + \frac{\lambda}{\mu}\pi_0 \cdot \sum_{i=1}^{\infty}\left(\frac{\lambda}{2\mu}\right)^{i-1} = 1$, $\lambda < 2\mu$ $\implies \pi_0 = \frac{2\mu-\lambda}{2\mu+\lambda}$

$\pi_i = (\cdots)\left(\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{2\mu}\right)^{i-1}$

Poisson Processes | For $(N_t)_{t\ge0} \sim$ PP(λ)

$\sim$ Poisson(λt)

Interarrival $S_1, S_2, \dots \overset{iid}{\sim}$ Exp(λ)

$T_n = \underbrace{\sum_{i=1}^{n} S_i}_{\sim\text{Erlang}}$

$N(t) = \max_{n\ge0}\{T_n \le t\}$

1.) $N_0 = 0$

2.) $N_t - N_s \sim \mathrm{Pois}(\lambda(t-s))$

3.) Independent Increments:

$(N_{t_1} - N_{t_0}) \cdots (N_{t_n} - N_{t_{n-1}})$ indep., $t_0 < t_1 < \cdots t_n$

$\mathbb{P}[N_t = n] = \mathbb{P}[t \in [T_n, T_{n+1})]$

$= \mathbb{E}[\mathbb{1}\{T_n \le t\}\mathbb{1}\{t < T_n + S_{n+1}\}]$

$= \int_0^t f_{T_n}(s)\,\mathbb{P}[S_{n+1} \ge t-s]\,ds$

$= \int_0^t \underbrace{f_{T_n}(s)}\, e^{-\lambda(t-s)}\,ds$

$\lambda \cdot e^{-\lambda s}(\lambda s)^{n-1}/(n-1)!$

Erlang: sum of n iid Exp

Uniform Arrival Times: Conditioned on $\{N_t = n\}$

$(T_1 \cdots T_n) \overset{d}{=} (U_{(1)} \cdots U_{(n)})$ w/ $U_{(i)}$ the ordered statistics of n indep. $U(0,t)$

Ex| Given $N_{120} = 10$.

$\mathbb{E}[N_{30}] = 10\cdot\frac{30}{120} = 10/4$, $N_{30} \sim \mathrm{Bin}(10, 30/120)$

Merging: $N_{1,t} + N_{2,t} \sim$ PP($\lambda_1 + \lambda_2$)

Splitting: $p_1 \cdots p_n > 0$, $\sum p_i = 1$. Input $(N_t)_{t\ge0} \sim$ PP(λ) and label i w.p. p_i s.t. $N_{t,i} \sim$ PP($p_i\lambda$)

↪ $f_{T_1 \cdots T_n}(t_1 \cdots t_n) = e^{-t_n}\prod_{i=1}^{n}(t_i - t_{i-1})$

Random Incidence Paradox: What's expected inter-arrival interval length that random time t_0 falls within?

$L = \underbrace{(T_{i+1} - t_0)}_{\sim\text{Exp}(\lambda)} + \underbrace{(t_0 - T_i)}$ | $\mathbb{P}[t_0 - T_i > s] = \mathbb{P}[\overbrace{N_{t_0} - N_{t_0-s} = 0}^{\text{no arrivals } \in (t_0-s,\, t_0)}]$

$= \exp(\lambda(t_0 - s - t_0))$

$= \exp(-\lambda s) \sim \mathrm{Exp}(\lambda)$

$\mathbb{E}[L] = \frac{1}{\lambda} + \frac{1-e^{-\lambda t}}{\lambda} \xrightarrow[t\to\infty]{} \frac{2}{\lambda}$

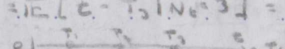
Ex| Students arrive from Ⓑ/Ⓢ $\sim$ PP(λ), PP(μ) indep.

Q1: $\mathbb{E}[T_n] = n/\lambda+\mu$

At time t, 3 students; $N_t = 3$

Q2: $\mathbb{E}$[time between last stu. & next stu. arrive]

$= \mathbb{E}[t - T_3 \mid N_t = 3] = t/4$



Q3: $N_t = 3$; $\mathbb{P}$[more Ⓑ than Ⓢ]

$= \mathbb{P}[\mathrm{Bin}(3, \lambda/\lambda+\mu) \in [2,3]]$

$= 3\lambda^2\mu + \lambda^3/(\lambda+\mu)^3$

Ex| Chance that

INFO

$H(X) = \sum_{x \in \mathcal{X}} p_X(x) \log \frac{1}{p_X(x)} = \log\left(\sum_{x\in\mathcal{X}} p_X(x) \frac{1}{p_X(x)}\right) = \log|\mathcal{X}|$

$= \mathbb{E}[\log 1/p_X(x)] \leq \log(\mathbb{E}(1/p_X(x)))$

$= -\mathbb{E}[\log p_X(x)]$

Expected surprise / uncertainty / information content

$\varphi(\mathbb{E}(Z)) \leq \mathbb{E}(\varphi(Z))$ where φ is a convex fctn

SOURCE CODING | $\forall \epsilon > 0$, $X_1 \dots X_n \overset{iid}{\sim} P_X$ can be losslessly compressed w/ $\leq n(H(X)+\epsilon)$ bits $\forall n >> 0$. Impossible for $< nH(X)$ bits

$H(X,Y) = \sum_{x,y} p_{XY}(x,y) \log 1/p_{XY}(x,y)$

$= \mathbb{E}[\log 1/p_{XY}(X,Y)]$

$= H(Y) + H(X|Y)$

Ex) BEC: $(X,Y) = \begin{cases} (0,0) & \text{w.p. } (1-p)/2 \\ (0,?) & p/2 \\ (1,?) & p/2 \\ (1,1) & (1-p)/2 \end{cases}$ $\quad H(X) = p\log\frac{1}{p} + q\log\frac{1}{q}$

$H(X,Y) = -2\frac{p}{2}\log(p/2) - 2\frac{(1-p)}{2}\log(1-p/2)$

$= 1 + (-p\log p - q \log q)$ for $q = 1-p$

$H(\text{Bern}(p))$

$H(Y) = p\log\frac{1}{p} + q\log\frac{1}{1-q}$ $= 1 - p + \underbrace{p\log\frac{1}{p} + q\log\frac{1}{q}}_{H(X)}$

Ex) X and Y indep:

$H(X|Y) = \sum_y p_Y(y) \sum_x p_X(x) \log 1/p_X(x)$

$= (\ 1\ ) H(X) = H(X)$

Conditional Entropy

$H(X|Y) = \sum_{y \in \mathcal{Y}} p_Y(y) \underbrace{\sum_{x \in \mathcal{X}} p_{X|Y}(x|y) \log 1/p_{X|Y}(x|y)}_{H(X|Y=y)}$

$H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$

Q3: WTS $H(X|Y) \leq p\log_2(n-1) + H(E)$ $\quad X \neq \hat{X} \Rightarrow |\mathcal{X}| \leq n-1$

$H(X|Y) \leq H(X|\hat{X})$ $\quad \underbrace{E=0 \Rightarrow X = \hat{X}}_{q H(X|\hat{X}, E=0) + p H(X|\hat{X}, E=1)}$

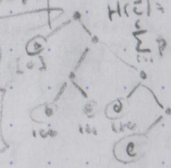
$= H(E, X|\hat{X}) = H(E|\hat{X}) + H(X|\hat{X}, E)$

$\leq H(E)$

Huffman

$L = H(\mathcal{C}) = \sum p_i \log \frac{1}{p_i}$

$\mathcal{C}$	$P(\mathcal{C})$
a	1/2
b	1/4
c	
d	
e	
f	1/16



MUTUAL INFO

$I(X;Y) := \sum_x \sum_y p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)}$

symmetric

$= \mathbb{E}[\ \cdot\ ]$

$= H(X) - H(X|Y)$

$= H(X) + H(Y) - H(X,Y)$

$C = \max_{P_X} I(X;Y)$ } fctn of $P_{Y|X}$

$R > C$: $P_e^{(n)} \to 1$

$R < C$: $\exists R$ s.t. $P_e^{(n)} < \epsilon$ $\forall n >> 1$

$I(X;Y|Z) = \mathbb{E}\left[\log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}\right]$

$= \mathbb{E}[-\log p(X|Z)] + \mathbb{E}[\log p(X|Y,Z)]$

$= H(X|Z) - H(X|Y,Z)$

$\sim \text{Unif}[1, 2^{nR}]$

Ex) Send L-bit msgs over BEC with rate $R = L/n$ w/ n-codewords

Q1: # Bits that can be reliably sent? $n(1-p)$ unerased bits, on avg.

Q2: Upperbound $\mathbb{P}[\text{Error}]$

$= \mathbb{P}\left[\bigcup_{i=2}^{2^L} \{c_1 = c_i\}\right] \leq \sum_{i=2}^{2^L} \mathbb{P}[c_1 = c_i]$ $\leq 2^L \cdot 1/2^{n(1-p)}$

$= 1/2^{n(1-p-R)}$

Ex) $X, Y \in [m]$; $\hat{X} = g(Y)$

$E = \mathbb{1}\{X \neq \hat{X}\}$, $p = \mathbb{P}[X \neq \hat{X}]$

Q1: WTS $0 = H(\hat{X}|Y) = \sum \sum p_{\hat{X}Y}(\hat{x},y) \log 1/p_{\hat{X}|Y}(\hat{x}|y) = 0$

$p_Y(y)\, \mathbb{1}\{\hat{x} = g(y)\}$

Q2: WTS $H(X|Y) \leq H(X|\hat{X})$

$I(X;\hat{X}|Y) = H(X|Y) - H(X|\hat{X},Y)$

$= H(\hat{X}|Y) - H(\hat{X}|X,Y) = 0$

$I(X;Y|\hat{X}) = H(X|\hat{X}) - \underbrace{H(X|\hat{X},Y)}_{H(X|Y)}$

$\geq 0 \Rightarrow H(X|\hat{X}) \geq H(X|Y)$

Times arrive $\sim PP(\lambda_f), PP(\lambda_m)$

Q1: $\mathbb{E}[T_2]$ w/ T_2 time w/ at least one male + female

$= \mathbb{E}[\text{First arrive}] + \mathbb{E}[T_f^* | m]\mathbb{P}[M] + \mathbb{E}[T_m^* | F]\mathbb{P}[F]$

$= (\lambda_f + \lambda_m)^{-1} + 1/\lambda_f \cdot \frac{\lambda_m}{\lambda_f+\lambda_m} + 1/\lambda_m \cdot \frac{\lambda_f}{\lambda_f+\lambda_m}$

Alt) $X_F \sim$ time for first female

$\mathbb{E}[\max(X_F, X_m)] = \mathbb{E}[\max(X_F,X_m) + \min(X_F,X_m)] - \mathbb{E}[\min(X_m, X_F)]$

$\mathbb{E}[X_F + X_m] = \frac{1}{\lambda_f} + \frac{1}{\lambda_m}$ $\quad 1/(\lambda_f + \lambda_m)$

Q2: $0 < a < b < c$

N_1 = # males in $[0,b]$

N_2 = # females in $[a,c]$

Whats dist $N_1 + N_2$? $\sim PP(\lambda_m(b) + \lambda_f(c-a))$

Q3: No females in $[0,1]$, four in $[1,2]$

$\mathbb{P}[\text{Exactly 2 of 4 are males}]$

$\hookrightarrow$ Treat male arrival as doubled

$P_{male} = 2\lambda_m / (2\lambda_m + \lambda_f) \Rightarrow \binom{4}{2}\left(\frac{2\lambda_m}{2\lambda_m+\lambda_f}\right)^2 \left(\frac{\lambda_f}{2\lambda_m+\lambda_f}\right)^2$

2 devices w/ lifetime $\overset{iid}{\sim} \text{Exp}(\mu)$ and service/repair $\sim \text{Exp}(\lambda)$

$\lambda \pi_0 = \mu \pi_1$, $\lambda \pi_1 = \mu \pi_2$ } $\pi_i = \left(\frac{\lambda}{\mu}\right)^i \pi_0$ $\quad \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu & -(\mu+\lambda) & \lambda \\ 0 & \mu & -\mu \end{bmatrix}$ $\quad 0 \underset{\mu}{\overset{\lambda}{\rightleftarrows}} 1 \underset{\mu}{\overset{\lambda}{\rightleftarrows}} 2$

$\sum \pi_i = 1 \Rightarrow \pi_0 = \frac{1}{1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2}$

2 Bartenders serve w/ rates 2 & 3 drinks/min and pour 1L/min. Choosing either randomly,

$\mathbb{E}[\text{volume}] = \mathbb{E}[V|B=1]\mathbb{P}[B=1] + \mathbb{E}[V|B=2]\mathbb{P}[B=2]$

$= \frac{1}{2}\left(\frac{2}{2} + \frac{2}{3}\right)L = 5/6\ L$

Kevin and Nathan shoot ball w/ rate $\lambda = 2$ ($2/3$, $1/3$)

Q1: $t = 3$: Kevin's expected lead

Each shot contributes $1/3 - 2/3$ towards the # shots $\sim PP(\lambda t)$ $\therefore -\frac{1}{3}(\lambda t) = -\frac{1}{3}(2)(3) = -2$

Q2: $t = 30$; $k = 7$. What's $\mathbb{E}[N | K = 5]$?

N over $[0, T_5] \sim PP(\frac{2}{3}\lambda T_5)$; $T_5 \sim 5^{th}$ of 7 ordered smpl from $U[0,t]$

$\therefore \mathbb{E}[\frac{2}{3}\lambda T] = \frac{2}{3} \cdot \frac{5}{7+1} \cdot \frac{2 \cdot 30}{1} = 25$

Cafe customers arrive rate λ

Q1: $\mathbb{E}[N(3,1) | N(0,2)] = \mathbb{E}[N(2,1)] + \frac{1}{2}N(0,2)$

Q2: Another store w/ rate μ.

$\mathbb{P}[\text{we get } k^{th} \text{ customer and comp. has } n]$

$= \binom{n+k-1}{k-1} \lambda^k \mu^n / (\lambda + \mu)^{n+k}$

$A_\epsilon^{(n)} := \{(x_1 \dots x_n) : p(x_1 \dots x_n) \geq 2^{-n(H(X)+\epsilon)}\} \subset \mathcal{X}^n$

1) $\mathbb{P}[\{X_1 \dots X_n\} \in A_\epsilon^{(n)}] \to 1$ w/ $n \to \infty$

2) $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$

First-Step

Let $A \subset S$ $\quad \mathcal{B}_i = \mathcal{B}_0 + (\mathcal{B}_1 - \mathcal{B}_0) + \dots + (\mathcal{B}_i - \mathcal{B}_{i-1})$

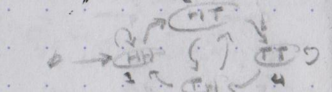
Hitting Time: $T_A = \min_{n \geq 0}\{X_n \in A\}$

Let $t_i := \mathbb{E}[T_A | X_0 = i] = \begin{cases} 1 + \sum_j P_{ij} t_j & \\ 0 & : i \in A \end{cases}$

$h_i = \frac{1}{q_i} + \sum_j \frac{P_{ij}}{q_i} h_j$

Ex) $\mathbb{E}[\text{flips til 2 tails occurs}]$ in a row

$P = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$



$t_{HH} = 1 + \frac{1}{2}(t_{HH} + t_{HT}) = 6$

$t_{HT} = 1 + \frac{1}{2}(t_{TH} + t_{TT}) = 4$

$t_{TH} = 1 + \frac{1}{2}(t_{HT} + t_{HH}) = 6$

$t_{TT} = 0$

Ex) Flip until 2 tails; $\mathbb{E}[\text{heads seen}]$

$t_S = \frac{1}{2}(t_H + t_T) = 3$ $\quad t_H = 1 + \frac{1}{2}(t_H + t_T) = 4$

$t_{TT} = 0$ $\quad t_T = \frac{1}{2}(t_H + t_{TT}) = 2$

Hitting Prob. $\alpha_i = \mathbb{P}[T_A < T_B | X_0 = i]$

$\alpha_i = \frac{1}{2}(\alpha_{i+1} + \alpha_{i-1})$ $\quad = \begin{cases} 0 & : i \in B \\ 1 & : i \in A \\ \sum_j P_{ij}\alpha_j & : o.w. \end{cases}$

$\alpha_{i+1} - \alpha_i = \alpha_i - \alpha_{i-1}$

Ex) Gambler's Ruin

$\mathbb{P}[T_R < T_0 | X_0 = i]$

$= \alpha_i = i/R$ as $\alpha_i = \frac{1}{2}\left(\frac{1}{R}(i-1) + (i+1)\right)$

Q) How many times visiting ①, starting there? $R = 100$

$\mathbb{E}_1[N_1]$

Let $\mathcal{B}_i = \mathbb{E}_i[N_1]$ st. $\mathcal{B}_1 = 1 + \frac{1}{2}\mathcal{B}_2 + \frac{1}{2}\mathcal{B}_0$

$\forall i \in [2, 99]$: $\mathcal{B}_i = \frac{1}{2}(\mathcal{B}_{i-1} + \mathcal{B}_{i+1})$

$\mathcal{B}_i - \mathcal{B}_{i+1} = \mathcal{B}_{i-1} - \mathcal{B}_i$

$\mathcal{B}_{99} = \frac{1}{2}\mathcal{B}_{98}$

$= \mathcal{B}_{98} - \mathcal{B}_{99} =: \Delta$ $\quad \mathcal{B}_K = \mathcal{B}_K - \overset{0}{\mathcal{B}_{100}}$

$= \sum_{i=K}^{99} \mathcal{B}_i - \mathcal{B}_{i+1}$

$\therefore \mathcal{B}_1 = 99\,\mathcal{B}_{99}$ $\quad = (100-K)\frac{\mathcal{B}_{99}}{\Delta}$

$= 1 + \frac{1}{2}\mathcal{B}_2 = 1 + \frac{98}{2}\mathcal{B}_{99}$

$\Rightarrow \mathcal{B}_{99} = 2/100$, $\mathcal{B}_1 = 99/50$

Say we've seen 2 cars & 3 trucks at t_0

Q) $\mathbb{E}[\text{time til } 10^{th} \text{ car seen}]$ $\quad$ 8 Exp(λ_c)

$= \mathbb{E}[T_{10}] = t_0 + 8\mathbb{E}[T_1]$

$= t_0 + 8/\lambda_c$

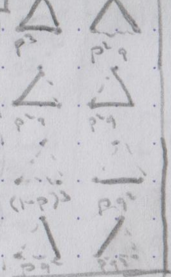
Q) @ T_{10}, $\mathbb{E}[\text{trucks seen}]$ $\quad = \lambda_T(t_0 + 8/\lambda_c)$

$= \mathbb{E}[\mathbb{E}[N | T_{10}]]$ $\quad - \lambda_T t_0$

$= \mathbb{E}[\lambda_T(T_{10} - t_0)] = 8\frac{\lambda_T}{\lambda_c} + 3$

@ t_0

Random Graphs) $n=3$



$G \sim G(n,p)$ has n nodes and $p\binom{n}{2}$ edges

$\hookrightarrow M$ edges: $p^M (1-p)^{\binom{n}{2}-M}$

Degrees $\sim$ Binomial:

$\mathbb{P}[\deg(v)=k]$
$= \binom{n-1}{k} p^k (1-p)^{n-1-k}$

Monotone property: Adding more edges preserves property

$\hookrightarrow$ If $A \subseteq B$, and P satisfies B, then P sat. A too.

Thresholds: $p \gg t_n \Rightarrow G$ has P w.h.p

$1/n^2$: (no) edges [Markov's]

$1/n$: (no) cycles

$\hookrightarrow p \ll 1/n \Rightarrow$ largest CC $O(\log n)$

$\hookrightarrow p \gg 1/n \Rightarrow$ " " $\Theta(n)$

Connectivity: For $\lambda > 0$, let $p_n = \lambda \frac{\log n}{n}$

$\lambda > 1$: $\mathbb{P}[G(n,p_n) \text{ connected}] \to 1$

$\lambda < 1$: " $\to 0$

PF) $\mathbb{P}[x=0] \le \text{Var}(X) / \mathbb{E}[X]^2$ Single Isolated Node

$\sum_i \text{Var}(I_i) + \sum_{i \ne j} \text{cov}(I_i, I_j)$ | $n \mathbb{E}[I_i]$
$n(1-p)^{n-1}$

$nq(1-q) + n(n-1)\frac{pq}{1-p}$

Ex) Cut) We create cut by select edge w.p. q.

WTF) $\mathbb{E}$[edges crossing cut]
$= \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} \cdot k(n-k) p$

condition on size k $\mathbb{E}$[crossing edges] given k-size cut

Ex) Each node has some $X_i \sim N(0,1)$ and edge $(i,j) \in E \Leftrightarrow X_i + X_j > c$

$\hookrightarrow$ Not Random Graph as knowing $(i,j) \in E$ increases $\mathbb{P}[(j,k) \in E]$.

WTF) $\mathbb{P}[X_i + X_j > c] = \mathbb{P}[N(0,1) < \frac{-c}{\sqrt{2}}]$
$N(0,2)$ $= \Phi(-c/\sqrt{2})$

JG) Gaus. R.V. $X = (X_1 \cdots X_n)^T$ has pdf.

$$f_X(x) = \frac{\exp(-\frac{1}{2}(x-\mu_X)^T \Sigma_X^{-1} (x-\mu_X))}{\sqrt{(2\pi)^n \det(\Sigma_X)}}$$

$X \sim N(\mu_X, \Sigma_X)$. For $\Sigma_X = \mathbb{E}[(x-\mu_X)(x-\mu_X)^T]$

Defns)

- Affine Transform of iid gaus r.v.

i.e. non-singular $\Sigma_X \Rightarrow X = \mu_X + AW$
$A \in \mathbb{R}^{n \times n}$, $W_i \overset{iid}{\sim} N(0,1)$

$\Rightarrow f_X(x) = f_W(A^{-1}(x-\mu_X)) / |\det(A)|$
$\exp(-\frac{1}{2}|A^{-1}(x-\mu_X)|) / (2\pi)^{n/2}$

$\cdots = \exp(-\frac{1}{2}(x-\mu_X)^T (AA^T)^{-1} (x-\mu_X)) \Big/ (2\pi)^{n/2} |\det(AA^T)|^{\frac{1}{2}}$

$\Rightarrow \Sigma_X = \mathbb{E}[AWW^TA^T]$
$= AA^T$

$\hookrightarrow \phi_X(t) = \mathbb{E}[\exp(it^TX)] = \phi_X(t_n)$

- 1D proj are gaus. R.V.s: $\forall a \in \mathbb{R}^n$

$a^T X \sim N(a^T \mu_X, a^T \Sigma_X a)$

$\text{Cov}(X,Y) = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}$ | $X_1 \cdots X_n$ pairwise uncorr $\Rightarrow \Sigma_X = \text{diag}(\sigma_1^2 \cdots \sigma_n^2)$

Amazing Fact: X, Y JG; always decompose:

$X = \mu_X + \Sigma_{XY} \Sigma_Y^{-1} (Y - \mu_Y) + V$

$V \sim N(0, \Sigma_X - \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX})$ indep of Y

$\Rightarrow \mathbb{E}[X|Y] = \mathbb{E}[\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y-\mu_Y) + V \mid Y]$
$= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y-\mu_Y)$
$= \mathbb{L}[X|Y]$

$\Rightarrow$ uncorrelated $\Leftrightarrow$ independence

Ex) $X, Y, Z \overset{iid}{\sim} N(0,1)$ | $\text{Cov}(X^2+Y^2, X+Y)$

Q1) $\mathbb{L}[X^2+Y^2 \mid X+Y] = ?$ $= \mathbb{E}[(\;)(\;)]$
$= \mathbb{E}[X^3 + X^2Y + Y^2X + Y^3] = 0$

$= \mathbb{E}[X^2+Y^2] = 0$ | $\text{Cov}((\cdot)(\cdot)) = \mathbb{E}[(\cdot)(\cdot)]$

Q2) $\mathbb{L}[X+2Y \mid X+3Y+4Z] = ?$ | $= \mathbb{E}[X^2] + 6\mathbb{E}[Y^2] = 7$

$\text{Var}(X+3Y+4Z) = V + 9V + 16V = 26$

$= \frac{7}{26}(X+3Y+4Z)$

Q3) $\mathbb{L}[(X+Y)^2 \mid X-Y]$ | $\text{cov}(X+Y, X-Y) = 0$
$\Rightarrow X+Y, X-Y$ indep

$= \mathbb{E}[(X+Y)^2]$

$= \text{Var}(X+Y) = 2$

Ex) $\text{Cov}(X,Y,Z) = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 3 & 2 \\ 4 & 2 & 3 \end{bmatrix}$ w/ $\mu_X = \mu_Z = 0$, $\mu_Y = 126$

$\mathbb{E}[Y|X,Z]$
$= a + bX + cZ$

$a = \mathbb{E}[\mathbb{E}[Y|X,Z]] = \mathbb{E}[Y] = \ldots$ | $\mathbb{E}[(Y-126-bX-cZ)Z] = \mathbb{E}[(Y-126)Z] - b\mathbb{E}[XZ] - c\mathbb{E}[Z^2] = 2-3c = 0$

$\mathbb{E}[(Y-126-bX-cZ)X] = \mathbb{E}[(Y-126)X] - b\mathbb{E}[X^2] - c\cancel{\mathbb{E}[ZX]}$ uncorr. $= 2 - 3b = 0 \Rightarrow b = 2/3$

$\Rightarrow c = 2/3$

Online) : Wlog $\mathbb{E}[X]=0$, $Y_1, Y_2, \cdots$ ortho.

Let $Y^n = (Y_1 \cdots Y_n)$, $\mu_i = \mathbb{E}[Y_i]$

$$\mathbb{L}[X|Y^{n+1}] = \mathbb{L}[X|Y^n] + \underbrace{\mathbb{L}[X|Y_{n+1}]}_{\frac{\text{Cov}(X, Y_{n+1})}{\text{Var}(Y_{n+1})}(Y_{n+1} - \mu_{n+1})}$$

PF) ortho principal:

$Y_k \perp (\mathbb{L}[X|Y^{n+1}] - X)$

$\Rightarrow 0 = \mathbb{E}[(\mathbb{L}[X|Y^n] + \mathbb{L}[X|Y_{n+1}] - X) Y_k]$

$\boxed{k=n+1}$ $= \underbrace{\mathbb{E}[(\mathbb{L}[X|Y_{n+1}] - X)Y_{n+1}]}_{\text{ortho. principal}} + \underbrace{\mathbb{E}[\mathbb{L}[X|Y^n]Y_{n+1}]}_{Y_i \perp Y_{n+1}}$

$\boxed{k \in [1,n]}$ $= \mathbb{E}[(\mathbb{L}[X|Y^n] - X)Y_k] + \mathbb{E}[\mathbb{L}[X|Y_{n+1}]Y_k]$

Kalman)

$X_{n+1} = AX_n + V_n \quad n \ge 0$
$Y_n = CX_n + W_n \quad n \ge 1$

$X_0, V_0, \cdots, W_0, \cdots$ uncorrelated

" " gaus. $\Rightarrow$ everything JG

Alg: Init $\hat{X}_{0|0} = 0$, $\Sigma_{0|0} = \text{Cov}(X_0)$

$\hat{X}_{n|n-1} = A\hat{X}_{n-1|n-1}$

$\tilde{Y}_n = Y_n - \underbrace{\mathbb{L}[Y_n | Y^{n-1}]}_{C\hat{X}_{n|n-1}}$

$\hat{X}_{n|n} = \hat{X}_{n|n-1} + K_n \tilde{Y}_n$
$= A\hat{X}_{n-1|n-1} + \underbrace{K_n}_{\Sigma_{n|n-1} C^T (C\Sigma_{n|n-1}C^T + \Sigma_W)^{-1}} (Y_n - CA\hat{X}_{n-1|n-1})$

$\Sigma_{n|n-1} = A\Sigma_{n-1|n-1}A^T + \Sigma_V$

$\Sigma_{n|n} = (I - K_nC)\Sigma_{n|n-1}$

Ex) $\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)$, $W = \begin{cases} 1 & Y > 0 \\ 0 & Y = 0 \\ -1 & Y < 0 \end{cases}$ $\uparrow$ sgn

Q1) $\mathbb{E}[WX|Y]$ JG
$= W\mathbb{E}[X|Y] = W\mathbb{L}[X|Y] = W(1 + \frac{1}{2}Y)$

Q2) $\mathbb{E}$ " $\neq \mathbb{L}[WX|Y]$ as $\mathbb{L}[\cdot]$ is linear
coef of Y varies w Fctn of Y

$\Rightarrow WX$ and Y not JG

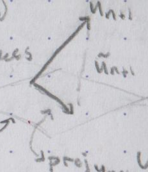
Ex) $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 & 3 & 1 \\ 3 & 9 & 3 \\ 1 & 3 & 2 \end{bmatrix}\right)$

$\mathbb{E}[X|Y,Z] = \Sigma_{X(Y,Z)} \Sigma_{(Y,Z)}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix}$

$= \begin{bmatrix} \mathbb{E}(XY) & \mathbb{E}(XZ) \end{bmatrix} \begin{bmatrix} \mathbb{E}Y^2 & \mathbb{E}ZY \\ \mathbb{E}YZ & \mathbb{E}Z^2 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix}$

$= \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 \\ 3 & 2 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix} = \frac{1}{3}Y$

Gram Schmidt) uncorrelated



Desire mutually orthogonal vecs st. $\text{span}\{u_1 \cdots v_n\} = \text{span}\{\tilde{u}_1 \cdots \tilde{u}_n\}$

$\tilde{Y}_{n+1} = Y_{n+1} - \mathbb{L}[Y_{n+1} | Y^n]$

$\Rightarrow \mathbb{L}[X|Y^n] = \mathbb{L}[X|\tilde{Y}^n]$

X, Y, Z zero-mean R.V.s; Y, Z uncorrelated

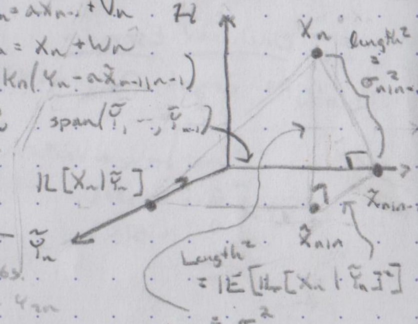
$\mathbb{L}[X|Y,Z] = \mathbb{L}[X|Y] + \mathbb{L}[X|Z]$

if correlated: $\tilde{Z} = \mathbb{L}[Z|Y]$
st $Y, \tilde{Z}$ uncorr, same span

$\therefore \mathbb{L}[X|Y,Z] = \mathbb{L}[X|Y] + \mathbb{L}[X \mid \underbrace{Z - \mathbb{L}[Z|Y]}_{\tilde{Z}}]$

Scalar)

$X_n = aX_{n-1} + V_n$
$Y_n = X_n + W_n$



$\hat{X}_{n|n} = a\hat{X}_{n-1|n-1} + K_n(Y_n - a\hat{X}_{n-1|n-1})$

$K_n = \sigma^2_{n|n-1} / \sigma^2_{n|n-1} + \sigma^2_W$

$\sigma^2_{n|n-1} = a^2 \sigma^2_{n-1|n-1} + \sigma^2_V$

$\sigma^2_{n|n} = (1-K)\sigma^2_{n|n-1}$

Ex) Even obs.

$Y_0, Y_2, \cdots Y_{2n}$

$\hat{X}_{2n|2n} = \mathbb{L}[X_{2n} | Y_0, Y_2 \cdots]$

$X_{2n+2} = aX_{2n+1} + V_{2n+1}$
$= a^2 X_{2n} + \underbrace{(aV_{2n} + V_{2n+1})}_{\sim N(0, (a^2+1)\sigma_V^2)}$

$\Rightarrow \hat{X}_{2n+2|2n+2} = a^2 \hat{X}_{2n|2n} + K_{2n+2}\tilde{Y}_{2n+2}$

$\tilde{Y}_{2n+2} = Y_{2n+2} - a^2 \hat{X}_{2n|2n}$

$K_{2n+2} = \sigma^2_{2n+2|2n} / \sigma^2_{2n+2|2n} + \sigma^2_W$

$\sigma^2_{2n+2|2n} = (a^2 \sigma^2_{2n|2n})^2 + (a^2+1)\sigma^2_V$

$\sigma^2_{2n+2|2n+2} = (1 - K_{2n+2})\sigma^2_{2n+2|2n}$

By lin. of LLSE: $\hat{X}_{2n+1|2n} = a\hat{X}_{2n|2n}$

Ex) $X \sim N(1,1)$, $Y \sim N(0,1)$ w/ $\text{cov}(X,Y) = \rho$

Q: $\mathbb{P}[X > Y]$?
$= \mathbb{P}[X - Y > 0]$

$\begin{cases} \mathbb{E}\, X - Y = 1 \\ \text{Var}\, X - Y = \text{Var}\, X + \text{Var}\, Y - 2\text{cov}(X,Y) = 2 - 2\rho \end{cases}$

$\therefore X - Y \sim N(1, 2-2\rho)$

$\Rightarrow \mathbb{P}[X > Y] = \mathbb{P}[N(0,1) > \frac{-1}{\sqrt{2-2\rho}}]$
$= 1 - \Phi(-1/\sqrt{2-2\rho}) = \Phi(\frac{1}{\sqrt{2-2\rho}})$

Inference $X \xrightarrow{\text{model}} P_{Y|X} \longrightarrow Y$

state of nature — observation

$\pi_i = \mathbb{P}[X=i]$

Maximum A Posterior: MAP(X|Y)

$\hat{X}_{MAP}(y) = \arg\max_x \underbrace{P_{X|Y}(x,y)}$

$\propto P_{Y|X}(y,x)\,\pi_x$

$= \begin{cases} 1 : L(y) > \pi_0/\pi_1 \\ 0 : L(y) < \pi_0/\pi_1 \end{cases}$

Maximum Likelihood Estimate:

$\hat{X}_{MLE}(y) = \arg\max_x P_{Y|X}(y|x)$

$= \hat{X}_{MAP}\big|_{\pi_x \sim \text{Uniform}}$

$= \begin{cases} 1 : L(y) > 1 \\ 0 : L(y) < 1 \end{cases}$

Binary Hyp. Testing:

$L(y) = \dfrac{P_{Y|X}(y|1)}{P_{Y|X}(y|0)} = \dfrac{P_{H_1}(y)}{P_{H_0}(y)}$

Null: $H_0 \Rightarrow Y \sim P_{Y|X=0}$

Alt: $H_1 \Rightarrow Y \sim P_{Y|X=1}$

Errors:

I: $\mathbb{P}[\hat{X}=1|X=0] = \alpha$; significance

↳ prob. false alarm

II: $\mathbb{P}[\hat{X}=0|X=1] = 1-\beta$; power

↳ prob. false detection

$\hat{X}_\beta = \arg\min_x \mathbb{P}[\text{II}] = \arg\max_x \mathbb{P}[\hat{X}=1|X=1]$ st $\mathbb{P}[\text{I}] \le \beta$

Neyman-Pearson:

$\hat{X}_\beta(y) = \begin{cases} 0 & L(y) < \lambda \\ \text{Bern}(\gamma) & L(y) = \lambda \\ 1 & L(y) > \lambda \end{cases}$

If $L(y)$ mono. increasing, let $L(y) = y$

↳ mono. decrease; flip ineq.

Ex $\mathbb{P}[X = \hat{X}_{MAP}] \le \mathbb{P}[X \ne \hat{X}]$

Let $\hat{X}_{MAP} = r(Y)$

$= \arg\max_x \mathbb{P}[X=x, Y=y]$

$= \arg\min_x \mathbb{P}[X \ne x, Y=y]$ ← $\hat{x}$ cond. indep. of $X|Y$

$\mathbb{P}[X \ne \hat{X}]$

$= \sum_y \mathbb{P}[X \ne \hat{X}, Y=y]$

$= \sum_y \sum_z \mathbb{P}[X \ne z, Y=y]\,\mathbb{P}[\hat{X}=z, Y=y]$

$\ge \sum_y \sum_z$ " $r(Y)$ "

$= \sum_y \mathbb{P}[X \ne r(Y), Y=y]$

$= \mathbb{P}[X \ne r(Y)]$

Alt as $X \in \{0,1\}$

$\hat{X}_{MAP} = \mathbb{1}\{P_{Y|X}(y|1)\pi_1 \ge P_{Y|X}(y|0)\pi_0\}$

$= \mathbb{1}\{L(y) \ge \pi_0/\pi_1\}$

$\mathbb{P}[X \ne \hat{X}]$

$= \pi_0 \mathbb{P}[\hat{X}=1|X=0] + \pi_1 \mathbb{P}[\hat{X}=0|X=1]$

$= \pi_0 \mathbb{E}[\hat{X}|X=0] + \pi_1(1 - \mathbb{E}[\hat{X}|X=1])$

$= \pi_1 \mathbb{E}[\frac{\pi_0}{\pi_1}\hat{X}|X=0] + 1 - \mathbb{E}[L(Y)\hat{X}|X=0]$

$= \pi_1 - \pi_1 \mathbb{E}[(L(Y) - \frac{\pi_0}{\pi_1})\hat{X}|X=0]$

$\le (\;\;"\;\;)\,\hat{X}_{MAP}$

Ex Type I $(\hat{X}=1|X=0)$ incur α

Type II $(\hat{X}=0|X=1)$ incur β

Want to min $\alpha\mathbb{P}[\text{I}] + \beta\mathbb{P}[\text{II}]$

$\alpha\pi_0\mathbb{P}[\text{I}] + \beta\pi_1\mathbb{P}[\text{II}]$

$= (\alpha\pi_0 + \beta\pi_1)\,\mathbb{P}[X \ne \hat{X}]$

$\Rightarrow$ min by MAP with scaled priors:

$\mathbb{P}[X=0] = \tilde{\pi}_0 = \alpha\pi_0 / (\alpha\pi_0 + \beta\pi_1)$

$\mathbb{P}[X=1] = \tilde{\pi}_1 = \beta\pi_1 / (\alpha\pi_0 + \beta\pi_1)$

Ex w $(x_1,y_1)\dots(x_n,y_n) = \vec{v}$

$Y = wx + \varepsilon,\ \varepsilon \sim \mathcal{N}(0,1)$

$L(\vec{v}|W=w) = \prod_1^n L((x_i,y_i)|W=w)$

$\propto \prod_1^n \exp(-(y_i - wx_i)^2/2\sigma^2)$

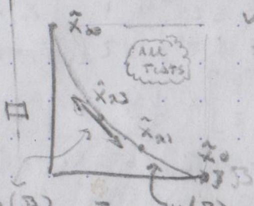
$\ell(\vec{v}|w) = \frac{-1}{2\sigma^2}\sum_1^n (y_i - wx_i)^2$] MLE

Ex $X_1 \dots X_n \overset{iid}{\sim} \text{Uni}[0,\phi]$

$L(\phi|X_1,\dots X_n) = \mathbb{1}\{0 \le X_1, \dots X_n \le \phi\}/\phi^n$

$\ell(\;\;"\;\;) = -\frac{n}{\phi}\mathbb{1}\{\;"\;\}$ → decreasing in ϕ for $\phi \ge X_1, \dots X_n$

$\therefore$ MLE $= X^{(n)} = \max\{X_1, \dots X_n\}$



$u(\beta) = \max_{\lambda \ge 0} \mathbb{P}[\hat{X}_\lambda = 0|X=1]$

$\lambda(\mathbb{P}[\hat{X}_\lambda = 1|X=0] - \beta)$

Ex $\Phi \sim U[0,100]$ & this $X \sim U[0,\Phi]$

Q1 MLE $[\Phi|X] = ?$

$\Phi \sim U \Rightarrow$ MLE $\equiv$ MAP

$\arg\max_\phi \frac{1}{\phi}\mathbb{1}[x \le \phi \le 100] = X$

Q2 $\mathbb{L}[\Phi|X] = \mathbb{E}[\Phi] + \dfrac{\text{cov}(\Phi,X)}{\text{Var}(X)}(X - \mathbb{E}X)$

50 $\quad \mathbb{E}[\mathbb{E}[X|\Phi]] = \mathbb{E}[\Phi/2] = 25$

$\text{cov}(\Phi,X) = \mathbb{E}[\Phi X] - \mathbb{E}[\Phi]\mathbb{E}[X]$

$\mathbb{E}[\mathbb{E}[\Phi X|\Phi]] = \mathbb{E}[\Phi^2/2]$ $\quad \mathbb{E}[\Phi^2/12]$ $\quad \text{Var}(\Phi/2)$

$\text{Var}(X) = \mathbb{E}[\text{Var}(X|\Phi)] + \text{Var}(\mathbb{E}[X|\Phi])$

$\mathbb{L}[X|Y] = \mathbb{E}[X] + \dfrac{\text{cov}(X,Y)}{\text{Var}(Y)}(Y - \mathbb{E}[Y])$

Ex $Y \sim \text{Bern}(p)$

$X = \begin{cases} 0 & p = 1/3 \\ 1 & p = 2/3 \end{cases}$

Q1 MLE $(X|Y)$

$= \mathbb{1}\{p(Y|1) > p(Y|0)\} = \mathbb{1}\{Y=1\}$

I: $\mathbb{P}[\hat{X}=1|X=0] = 1/3$

II: $\mathbb{P}[\hat{X}=0|X=1] = (1 - \frac{2}{3}) = 1/3$

Q2 Plot $\begin{cases} \hat{X}_{\lambda,1} = \mathbb{1}[L(y) \ge \lambda] \text{ w.p. } \gamma \\ \hat{X}_{\lambda,0} = \mathbb{1}[L(y) > \lambda] \text{ w.p. } 1-\gamma \end{cases}$

$L(y) = \begin{cases} 1/2 & y=0 \\ 2 & y=1 \end{cases}$

$\therefore \hat{X}_\lambda = \begin{cases} 1 & \lambda < 1/2 \\ Y & \lambda \in (1/2, 2] \\ 0 & \lambda > 2 \end{cases}$

Q3 min II for $\beta = 1/2 \ge$ PFA

$\Rightarrow$ Take Y w.p. $3/4$

1 w.p. $1/4$ $\quad (\frac{1}{4}, \frac{1}{2})$

Ex Flip till heads; $0 < p < q < 1$

$X = \begin{cases} 0 : \text{bias } p \\ 1 : \text{bias } q \end{cases}$

$L(y) = \dfrac{p(Y|1)}{p(Y|0)} = \dfrac{(1-q)^{y-1}q}{(1-p)^{y-1}p}$

$p < q \Rightarrow$ strict. decrease

$\therefore L(y) \sim \mathbb{1}\{Y < t\}$

$\mathbb{P}[Y < t|X=0] = 1 - (1-p)^{t-1} \le \beta$

$\therefore t = 1 + \lfloor \log 1-\beta / \log 1-p \rfloor$

$\mathbb{P}[\hat{X}=1|X=0] = $ PFA

$= \mathbb{P}[Y < t|0] + \gamma\mathbb{P}[Y = t|X=0]$

$= 1 - (1-p)^{t-1} + \gamma p(1-p)^{t-1} = \beta \Rightarrow \gamma = \beta - 1 + (1-p)^{t-1} / p(1-p)^{t-1}$

Estimation $Y \xrightarrow[\text{estimate procedure}]{} \hat{X}(Y)$ (observation → estimate)

MMSE: $\arg\min_{\hat{x}} \mathbb{E}[(X - \hat{X}(Y))^2]$

$= \mathbb{E}[X|Y]$

↳ we don't often know P_{XY} exactly, and $\mathbb{E}[X|Y]$ intractable

Error

$\mathbb{E}[X - \mathbb{E}[X|Y]] = \text{Var}(X) - \text{Var}(\mathbb{E}[X|Y])$

LLSE: $\hat{X}(Y) \approx \alpha + \sum_{i=1}^n b_i Y_i$

$\mathbb{L}[X|Y] = \arg\min_{\hat{x} \text{ linear}} \mathbb{E}[(X - \hat{X}(Y))^2]$

$= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y)$

w/ $\begin{cases} \Sigma_{XY} = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)^T] \\ \Sigma_Y = \mathbb{E}[(Y-\mu_Y)(Y-\mu_Y)^T] \\ \mu_X = \mathbb{E}[X],\ \mu_Y = [\mathbb{E}[Y_1] \dots]^T \end{cases}$

↳ 2nd order stats from data

Vector-case: $X = [X_1 \dots X_k]^T$

$\mathbb{L}[X|Y] = \sum_{i=1}^k \min_{\hat{x}_i} \mathbb{E}[(X_i - \hat{X}_i(Y))^2]$

Quad: $\hat{X}(Y) = a + b_1 Y + b_2 Y^2$

Linear in a, b_1, b_2; $\tilde{Y} = [Y \;\; Y^2]^T$

st $\mathbb{L}[X|\tilde{Y}] = \Sigma_{X\tilde{Y}}\Sigma_{\tilde{Y}}^{-1}\tilde{Y}$

↖ zero-mean

Error squared:

$\mathbb{E}[(X - \mathbb{L}[X|Y])^2]$

$= \text{Var}(X) - \dfrac{\text{cov}(X,Y)^2}{\text{Var}(Y)}$

Linear Regression: $Y = AX + Z$

w/ $\Sigma_X = \sigma_X^2 I$, $\Sigma_Z = \sigma_Z^2 I$ uncorrelated; $A \in \mathbb{R}^{n \times k}$, $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^k$; zero mean

$\mathbb{L}[X|Y] = \Sigma_{XY}\Sigma_Y^{-1}Y$

$= (\mathbb{E}[XY^T])(\mathbb{E}[YY^T])^{-1}Y$

$\sigma_X^2 A^T \quad \mathbb{E}[(AX+Z)(X^TA^T + Z^T)]$

$= \sigma_X^2 AA^T + \sigma_Z^2 I$

$= \sigma_X^2 A^T(\sigma_X^2 AA^T + \sigma_Z^2 I)^{-1}Y$

$= A^T\left(AA^T + \dfrac{\sigma_Z^2}{\sigma_X^2}I\right)^{-1}Y$

↖ $\sigma_X^2 \to \infty$ if unknown (non-Bayes)

Hilbert Spaces

$\langle X, Y \rangle := \mathbb{E}[XY]$

↳ $\|X\|^2 = \langle X, X \rangle = \mathbb{E}[X^2]$

Properties of LLSE:

$\langle \mathbb{L}[X|Y] - X, u \rangle = 0$

$\mathbb{E}[(\mathbb{L}[X|Y] - X)u]$

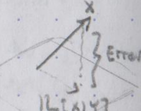
$= \mathbb{E}[(\;\;"\;\;)(a + \sum b_i Y_i)]$

$\Rightarrow \mathbb{E}[\mathbb{L}[X|Y]] = \mathbb{E}[X]$ } unbias est.

$\mathbb{E}[\mathbb{L}[X|Y]Y_i] = \mathbb{E}[XY_i]$ } ortho. principl

$\therefore \mathbb{E}[\mathbb{L}[X|Y]Y^T] = \mathbb{E}[XY^T]$

Error uncorrelated $\Longleftrightarrow$ Est $\perp$ observ



$\mathbb{E}[(X - \mathbb{L}[X|Y])^2]$

$= \|X - \mathbb{L}[X|Y]\|^2$

$= \mathbb{E}[X^2] - \mathbb{E}[\mathbb{L}[X|Y]^2]$

$= \text{Var}(X) - \mathbb{E}[(\Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y))^2]$

$= \text{Var}(X) - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$

$X \perp 1 \Rightarrow 0 = \mathbb{E}[X \cdot 1]$ } 0-mean

$\|X\| = \sqrt{\mathbb{E}[X^2]} = \sigma_X = \sqrt{\text{Var}(X)}$

iff $\mathbb{E}[X^2] = \text{Var}(X)$; $\mathbb{E}X = 0$

$\mathbb{E}[X] = \langle X, 1 \rangle = \arg\min_{x \in \mathbb{R}} \mathbb{E}[(X - x)^2]$