Uniform(a,b)	$b-a+1$	$\frac{}{2}$	$\frac{}{12}$	
Bernoulli(p)	$\begin{cases} p & \text{if } k=1 \\ (1-p) & k=0 \end{cases}$	p	$p(1-p)$	
Binomial(n,p)	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	biased coin tossed n times, indep of other tosses. X is # heads in n tosses
Geo(p)	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$	repeatedly toss biased coin X := # tosses until 1st H
Pois(λ)	$\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ	probability of a given # (k) events occurring in a fixed interval of time/space, with a constant mean rate ($\mathbb{E}[X]=\lambda$) and indep of time since last event.

take: $P_X(x) = P\{X=x\}$

$\sum_x P_X(x) = 1$

Distribution	PDF $f_X(x)$	CDF $F_X(x)=P[X\le x]$	$\mathbb{E}[X]$	Var(x)
Uniform(a,b)	$\begin{cases} \frac{1}{b-a} & a\le x\le b \\ 0 & o/w \end{cases}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exp(λ) "memoryless"	$\begin{cases} \lambda e^{-\lambda x} & \text{if } x\ge 0 \\ 0 & o/w \end{cases}$	$1-e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal/Gaussian ($\mathcal{N}(\mu,\sigma^2)$)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\Phi(x)$	μ	σ^2

Standard Normal CDF:

$\Phi(y) = P[Y\le y] = P[Y<y] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left\{-\frac{t^2}{2}\right\} dt$] Normal RV with 0 mean and unit variance

↳ CDF calculation of $X \sim \mathcal{N}(\mu,\sigma^2)$

$Y := \frac{X-\mu}{\sigma}$ $(Var(Y)=1, \mathbb{E}[Y]=0)$

$\mathbb{P}[X\le x] = \mathbb{P}\left[\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right] = \mathbb{P}\left[Y\le \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Probability Axioms (Kolmogorov Axioms)

1) $P(A)\ge 0 \quad \forall A \in F$
2) $P(A\cup B) = P(A)+P(B)$ if A, B disjoint
3) $P(\Omega)=1$

Probability Law Properties

1) if $A\subset B$, $P(A)\le P(B)$
2) $P(A\cup B) = P(A)+P(B)-P[A\cap B]$
3) $P(A\cup B)\le P(A)+P(B)$
4) $P(A\cup B\cup C) = P(A)+P(A^c\cap B)+P(A^c\cap B^c\cap C)$
5) $P(A\cup B) = P(A)+P(A^c\cap B)$
6) $P(B) = P(A\cap B)+P(A^c\cap B)$

Conditional Probability

$P(A|B) = \frac{P(A\cap B)}{P(B)}$

↳ multiplication rule:

$P(\bigcap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots P(A_n|\bigcap_{i=1}^{n-1} A_i)$

↳ total probability thm:

$P(B) = P(A_1\cap B)+\dots+P(A_n\cap B)$

$= P(A_1)P(B|A_1)+\dots+P(A_n)P(B|A_n)$

↳ Bayes' rule

$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$

$= \frac{P(A_i)P(B|A_i)}{P(A_i)P(B|A_i)+P(A_i^c)P(B|A_i^c)}$

Independence

$P(A|B) = P(A)$

$P(A\cap B) = P(A)P(B)$

↳ conditional independence:

$P(A\cap B|C) = P(A|C)P(B|C)$

$P(A|B\cap C) = P(A|C)$

Counting

- Principle: if there's a sequence of indep. events that can occur $a_1,\dots,a_n$ ways, the # of ways ∀ events to occur is $\prod_{i=1}^n a_i$
- # subsets of an n-element set: 2^n
- k-permutations: n distinct objects $k\le n$ - # ways we can pick k out of the n (ie # of k-length permutations): $\frac{n!}{(n-k)!}$
- combinations: (no ordering); # of k-element subsets of a given n-element set: $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$
- partitions: n objects, $\sum_i n_i = n$, r disjoint groups w/ i^{th} group containing n_i items:

$\binom{n}{n_1}\binom{n-n_1}{\,}\cdots\binom{n-\dots-n_{r-1}}{\,} = \frac{n!}{} \dots \binom{n}{\,}$

Variance

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

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

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

$var(\alpha X+\beta) = \alpha^2 Var(X)$

- Law of Conditional Variance:

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

- independent rv's:

$Var(X+Y) = Var(X)+Var(Y)$

Standard deviation(σ)

$\sigma_X = \sqrt{Var(X)}$

Covariance & Correlation

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

↳ if $Cov(X,Y)=0$, X & Y are uncorrelated

↳ if X, Y indep, $cov(X,Y)=0$

- correlation coeff (ρ):

$\rho = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}}$

Expectation $\mathbb{E}[X]$

- mean/weighted average

$\mathbb{E}[X] = \sum_x x P_X(x)$

- n^{th} moment of an rv:

$\mathbb{E}[X^n] = \sum_x x^n P_X(x)$

- LOTUS:

$\mathbb{E}[g(X)] = \sum_x g(x)P_X(x)$

- total expectation thm:

$\mathbb{E}[X] = \sum_y P_Y(y)\mathbb{E}[X|Y=y]$

↳ Conditional Expectation:

$\mathbb{E}[X|Y] = \sum_x x P_{X|Y}(x|y)$

↳ Law of iterated $\mathbb{E}$:

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

↳ Tower Property

$\mathbb{E}[X] = \sum_i \mathbb{E}[X|A_i]P[A_i]$

↳ if A_i a countable partition of Ω

↳ Linearity of $\mathbb{E}$:

$\mathbb{E}[\alpha X+Y] = \alpha\mathbb{E}[X]+\mathbb{E}[Y]$

- independent rvs (X, Y)

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

- multiple RV's $\mathbb{E}$:

$\mathbb{E}(g(X,Y)) = \sum_{x,y} g(x,y)P_{X,Y}(x,y)$

Cumulative Distribution Functions (CDFs)

$F_X(x) = P[X\le x]$

↳ Properties:

1. Monotonically non-decreasing $(x\le y \Rightarrow F_X(x)\le F_X(y))$
2. $F_X(x)\to 0$ as $x\to-\infty$; $F_X(x)\to 1$ as $x\to\infty$
3. Discrete case → PMF & CDF can be obtained from each other by summing/differencing:

$F_X(k) = \sum_{i=-\infty}^{k} P_X(i)$

$P_X(k) = P[X\le k]-P[X\le k-1]$

$= F_X(k)-F_X(k-1)$

4. Continuous case → PDF & CDF obtained by integrating/differentiating

$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

$f_X(x) = \frac{d F_X(x)}{dx}$

5. Continuous from the right

Order statistics

- $X_1,\dots,X_n$ iid rv's, sorted s.t:

$X^{(1)}\le X^{(2)}\le\dots\le X^{(n)}$

↳ $X^{(1)} = \min\{X_1(\omega),\dots,X_n(\omega)\}$

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

- if X_i's cts rv's w/ density f_X then:

$f_{X^{(i)}}(x) = n\binom{n-1}{i-1}F_X(x)^{(i-1)}(1-F_X(x))^{n-i}f_X(x)$

$\approx \delta P\{X^{(i)}\in(x,x+\delta)\}$

$= P(X=x, Y=y)$

- using joint PMF to calc. marginal PMF:

$P_X(x) = \sum_y P_{X,Y}(x,y)$

→ Conditional PMFs

$P_{X|A} = P(X=x|A)$

$= \frac{P(\{X=x\}\cap A)}{P(A)}$

$P_{X|Y} = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

- using conditional PMF to get joint PMF:

$P_{X,Y}(x,y) = P_Y(y)P_{X|Y}(x|y)$

$= P_X(x)P_{Y|X}(y|x)$

- independence:

$P_{X,Y}(x,y) = P_X(x)P_Y(y)$

$P_{X|Y}(x|y) = P_X(x)$

↳ conditional independence:

$P(X=x,Y=y|A) = P_{X|A}(x)P_{Y|A}(y)$

$P_{X|Y,A}(x) = P_{X|A}(x)$

Expectation (Continuous)

$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$

$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$

$\mathbb{E}[g(X)|A] = \int_{-\infty}^{\infty} g(x)f_{X|A}(x)\,dx$

$\mathbb{E}[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)\,dx$

$\mathbb{E}[g(X,Y)|Y=y] = \int_{-\infty}^{\infty} g(x,y)f_{X|Y}(x|y)\,dx$

Total Expectation Thm:

$\mathbb{E}[X] = \int \mathbb{E}[X|Y=y]f_Y(y)\,dy$

$\mathbb{E}[g(X,Y)] = \int \mathbb{E}[g(X,Y)|Y=y]f_Y(y)\,dy$

Moment Generating Fcns (MGFs)

$M_X(s) = \mathbb{E}[e^{sX}]$

- For discrete rvs; MGF given by:

$M(s) = \sum_x e^{sx}P_X(x)$

- For continuous rvs:

$M(s) = \int_{-\infty}^{\infty} e^{sx}f_X(x)\,dx$

- Sum of indep rv's:

$W := X+Y$

$M_W(s) = \mathbb{E}[e^{s(X+Y)}]$

$= M_X(s)M_Y(s)$

- Properties:

1. $M_X(0)=1$
2. $\frac{d}{ds}M_X(s)\Big|_{s=0} = \mathbb{E}[X]$
3. $\frac{d^n}{ds^n}M_X(s)\Big|_{s=0} = \mathbb{E}[X^n]$
4. if $Y=\alpha X+\beta$: $M_Y(s) = e^{sb}M_X(as)$

- Transforms for common Discrete RVs:
  - Bernoulli(p): $M_X(s) = 1-p+pe^s$
  - Binomial(n,p): $M_X(s) = (1-p+pe^s)^n$
  - Geometric(p): $M_X(s) = \frac{pe^s}{1-(1-p)e^s}$
  - Poisson(λ): $M_X(s) = e^{\lambda(e^s-1)}$
  - Uniform(a,b): $\frac{e^{as}\quad e^{(b-a+1)s}\quad -1}{}$

↳ $f_X(x)\ge 0$

$\int_{-\infty}^{\infty} f_X(x) = 1$

- Jointly continuous PDFs:

$P[a\le X\le b, c\le Y\le d] = \int_c^d\int_a^b f_{X,Y}(x,y)\,dx\,dy$

↳ using joint to get marginal PDF:

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

- conditional PDFs:

$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

$P[X\in A|Y=y] = \int_A f_{X|Y}(x|y)\,dx$

- independence:

$f_{X,Y}(x,y) = f_X(x)f_Y(y)$

- Linear fcns of rv's:

$Y := aX+b$

$f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$

- Monotonic fcns of CRVs:

g := monotonic

$y = g(x)$ if $x = h(y)$

↳ if h has 1st derivative:

$f_Y(y) = f_X(h(y))\left|\frac{dh(y)}{dy}\right|$

- A a subset of the real line, $P(X\in A)>0$ then:

$f_{X|A}(x) = \begin{cases} f_X(x)/P(X\in A) & \text{if } x\in A \\ 0 & o/w \end{cases}$

&

$P[X\in B|X\in A] = \int_B f_{X|A}(x)\,dx$

- inference/Continuous Bayes

$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{f_Y(y)} = \frac{f_X(x)f_{Y|X}(y|x)}{\int f_X(t)f_{Y|X}(y|t)\,dt}$

- Continuous convolutions

$F_W(\omega) = P[W\le\omega]$

$= P[X+Y\le\omega]$

$= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\omega-x} f_X(x)f_Y(y)\,dy\,dx$

$= \int_{x=-\infty}^{\infty} f_X(x)F_Y(\omega-x)$

$f_W(\omega) = \int_{x=-\infty}^{\infty} f_X(x)f_Y(\omega-x)\,dx$

- Convolutions $(p_W(\omega))$

$W = X+Y$

$p_W(\omega) = P[X+Y=\omega]$

$= \sum_x P[X=x\cap Y=\omega-x]$

$= \sum_x p_X(x)p_Y(\omega-x)$

convolution of PMFs of X & Y.

- Transforms for common Continuous RVs:
  - Uniform(a,b): $M_X(s) = \frac{1}{b-a}\cdot\frac{e^{sb}-e^{sa}}{s}$
  - Exponential(λ): $M_X(s) = \frac{\lambda}{\lambda-s}$
  - Normal(μ, σ²): $M_X(s) = e^{[(\sigma^2 s^2/2)+\mu s]}$

$\mathbb{E}[X] = \sum_{k=1}^{\infty} P[X\ge$

entropy:

$H(X) := \sum_x P_X(x)\log$

$= \mathbb{E}[\log(\frac{1}{P_X(x)}$

Tower Property

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

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

Definition of being a σ-algebra:

1. Nonempty (contains Ω itself)
2. Closed under complement: $A\in F \Rightarrow A^c\in$
3. Closed under countable unions: $A_1,\dots,A_n\in$ $\Rightarrow \bigcup_{i=1} A_i\in F$

Derangements

- permutation where no element in a set is in its og pos

$D_n = !n = n!\sum_{k=0}^{n}$

$\sum_{i=1}^{n} i^{-1} \approx \ln$

$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}$

Pairwise indep

$P(A_i\cap A_j) = P(A_i$

Mutual indep:

$P(\bigcap_{i=1}^n A_i\cap B) = \prod_{i=1}^n P($

Minimum of 2 Geometrics is Geometric: $X\sim Geo(p)$ $Y\sim Geo(q)$

$Z = \min(X,Y)$

$P(Z>k) = P(X>k, Y>k) = P(X>k)P(Y>k)$

$= (1-p)^k(1-q)^k$

$= [(1-p)(1-q)]^k$

$Z\sim Geo(1-(1-p)(1-q))$

Lecture 3 Passcode: Teapot

Lecture 5 Passcode: Mirror

Lecture 8 Passcode: Lewis

Lecture ? PSC: Shannon

- expected time until you hit a certain value
- Let $A \subset S$ & define hitting time as

  $T_A = \min\{n \geq 0 : X_n \in A\}$
- Strategy: Define $h(i) := \mathbb{E}[T_A \mid X_0 = i]$

  $h(i) = 0 \quad i \in A$

  $h(i) = 1 + \sum_j P_{ij} h(j) \quad i \notin A$

Hitting Probabilities

- probability of hitting state A before state B
- For $a, b \in S$ define

  $T_a = \min\{n \geq 0 : X_n = a\}$

  $T_b = \min\{n \geq 0 : X_n = b\}$
- Strategy: Define $h(i) := P\{T_a < T_b \mid X_0 = i\}$

  → observe:

  $h(a) = 1 \quad h(b) = 0$

  $h(i) = \sum_j P_{ij} h(j) \quad i \notin \{a, b\}$

Poisson Processes $PP(\lambda)$

- Poisson process w rate λ is a counting process with iid $Exp(\lambda)$ interarrival times
- Formal def:
  - $S_1, S_2, \dots \overset{iid}{\sim} Exp(\lambda)$ $(\lambda > 0)$ are sample interarrival times
  - For each $n \geq 1$ define

    $T_n = \sum_{j=1}^{n} S_j$
  - The fcn $N(t)$ represents number of arrivals at time t

    $N(t) = \max\{n \geq 0 : T_n \leq t\}$
  - The sequence $\{N(t)\}_{t \geq 0}$ is a $PP(\lambda)$
- $N(t_1, t_2) := N(t_2) - N(t_1)$ (# of arrivals in an interval $[t_1, t_2]$)

Properties of a PP $\{N(t)\}_{t \geq 0} \sim PP(\lambda)$

1. Stationary increments:

   For every $t, s > 0$,

   $N(t, t+s) \overset{d}{=} N(s)$

   ie, $N(t, t+s)$ has the same distribution as $N(s)$
2. independent increments:

   For $0 < t_1 < \dots < t_k$ the set of rv's $N(t_1), N(t_1, t_2), \dots, N(t_{k-1}, t_k)$ are jointly independent
3. $N(t) \sim Poisson(\lambda t)$

   ↳ ie, the # of arrivals are Poisson

   $$P\{N(t) = n\} = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad n \geq 0$$

Erlang Distribution Erlang$(n; \lambda)$

- used to model the time between n ~~independent~~, identical events that occur at an average rate λ.

  ↳ $T_n :=$ n^{th} arrival time for a poisson process w/param λ, then:

  $$f_{T_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$$

Merging

$N \sim PP(\lambda)$ $M \sim PP(\mu)$, indep of N

⇒ $N + M \sim PP(\lambda + \mu)$

Splitting

$N \sim PP(\lambda)$

$B_1, B_2, \dots \overset{iid}{\sim} Bernoulli(p)$, indep of N

$N_0(t) = |\{i : B_i = 0, i \leq N(t)\}|$

$N_1(t) = |\{i : B_i = 1, i \leq N(t)\}|$

⇒ $N_0(t) \sim PP(\lambda p)$

$N_1(t) \sim PP(\lambda(1-p))$

& N_0, N_1 indep of each other

$(T_1, \dots, T_n) \mid \{N_t = n\}$ = order statistics on n iid $Unif(0,t)$ rvs.

Random Incidence Paradox

Let $N_t \sim PP(\lambda)$. Suppose I pick a time to far into the future, the expected time between the previous & next arrival is $\frac{2}{\lambda}$

Memorylessness of PP

$$P[T > t+s \mid T > t] = \frac{P[T > t+s, T > t]}{P[T > t]}$$

$$= \frac{P[T > t+s]}{P[T > t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s}$$

Kth arrival time of PP

- $T_n \sim Erlang(n; \lambda)$

  $\mathbb{E}[T_n] = \frac{n}{\lambda}$

  $var(T_n) = \frac{n}{\lambda^2}$

Arrival of Merged Processes

- has probability $\frac{\lambda_1}{\lambda_1 + \lambda_2}$ of originating from the 1st process and probability $\frac{\lambda_2}{\lambda_1 + \lambda_2}$ of originating from the 2nd process

Continuous Time MCs

- characterized by rate matrix/generator matrix Q
- Q-matrix has 3 properties:
  1. off diagonal elements are non-negative

     $[Q]_{ij} \geq 0 \quad \forall i \neq j \in S$
  2. rows sum to 0

     $\sum_j [Q]_{ij} = 0 \quad \forall i \in S$

     equivalently:

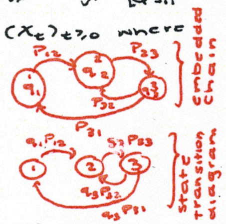
     $[Q]_{ii} = -\sum_{i \neq j} [Q]_{ij}$
- transition rate for state i (q_{ii}) tells you exact distr. for holding state i

  $q_{ii} = -[Q]_{ii}$
- transition probabilities (P_{ij}) are defined for the embedded chain/jump chain as:

  $$P_{ij} = \frac{[Q]_{ij}}{q_i} \geq 0 \quad \Big\} \quad \sum_{j \neq i} P_{ij} = \sum_{j \neq i} \frac{[Q]_{ij}}{-[Q]_{ii}} = 1$$
- CTMC is process $(X_t)_{t \geq 0}$ where X_t = state at time t

$$Q = \begin{bmatrix} -q_1 & q_1 P_{12} & 0 \\ 0 & -q_2 & q_2 P_{23} \\ q_3 P_{31} & q_3 P_{32} & -q_3 \end{bmatrix}$$

↳ conventionally, we draw state transition w/arrows labeled by transition rates →



Stationary Distribution & Hitting Times of CTMCs

- Stationary Distribution satisfies

  $\pi Q = 0 \quad \sum_i \pi_i = 1$

  ↳ ie:

  $$\forall i \quad \underbrace{\pi_i(-Q_{ii})}_{\text{rate out}} = \underbrace{\sum_{j \neq i} \lambda_{j \to i} \pi_j}_{\text{rate in}}$$

Remember the [illegible]

→ male & female titans arrive indep according to $PP(\lambda_m)$ & $PP(\lambda_F)$

1) $\mathbb{E}$[time until 1st titan arrives]

   $= \frac{1}{\lambda_F + \lambda_m} = \mathbb{E}[T_1]$

2) $\mathbb{E}$[time until 1st female & 1st male arrive]

   $= \mathbb{E}[T_1] + P[M]\,\mathbb{E}[T_1^F \mid M] + P[F]\,\mathbb{E}[T_1^M \mid F]$

   $= \frac{1}{\lambda_F + \lambda_m} + \frac{1}{\lambda_F} \cdot \frac{\lambda_m}{\lambda_F + \lambda_m} + \frac{1}{\lambda_m} \cdot \frac{\lambda_F}{\lambda_F + \lambda_m}$

3) $0 < a < b < c$. N_1 = # males arriving during $[0, b]$. N_2 = # females during $[a, c]$.

   $N_1 + N_2 \sim Pois(\lambda_m b + \lambda_F (c - a))$

4) No females during $[0, 1]$. 4 titans arrive in $[0, 2]$, P[exactly 2 titans are male] = ?

   ↳ # males in $[0, 2]$ is # arrivals of a $PP(2\lambda_m)$ in $\frac{1}{2}$ the interval ($[1, 2]$). merge this w/ female arrival process.

   P[male titan]: $\frac{2\lambda_m}{2\lambda_m + \lambda_F}$

   $$P[\text{exactly 2 T's M}] = \binom{4}{2} \left(\frac{2\lambda_m}{2\lambda_m + \lambda_F}\right)^2 \left(\frac{\lambda_F}{2\lambda_m + \lambda_F}\right)^2$$

5) At time t, see 3 titans. What's expected time btwn last titan arrival & next titan after t?

   $\mathbb{E}[t - T_3 \mid N(t) = 3] = \frac{1}{4} t$

   ↳ next titan takes $\frac{1}{\lambda_F + \lambda_m}$ time in $\mathbb{E}$,

   so, we get that

   $\mathbb{E}[\text{time btwn}] = \frac{1}{4t} + \frac{1}{\lambda_F + \lambda_m}$

6) P[more F than M at time t] = ?

   $p_F = \frac{\lambda_F}{\lambda_F + \lambda_m}$

   ⇒ # titans in 1st 3 arrivals distributed as $Bin(3, p_F)$

   ⇒ $P[\text{more } F] = \frac{3\lambda_F^2 \lambda_m + \lambda_F^3}{(\lambda_F + \lambda_m)^3}$

7) At time t, 2 F, 3M. What's $\mathbb{E}$[# M when it has 10 F]?

   $S :=$ # males btwn t and T_{10}

   $\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S \mid T_{10}]]$

   $S \sim Pois(\lambda_m (T_{10} - t))$

   $\mathbb{E}[S] = \mathbb{E}[\lambda_m (T_{10} - t)] = \lambda_m (\mathbb{E}[T_{10}] - t)$

   $= \lambda_m \frac{(10 - 2)}{\lambda_F} = \frac{\lambda_m}{\lambda_F} 8$

   ↳ $\mathbb{E}[\text{Males at } T_{10}] = 3 + \frac{\lambda_m}{\lambda_F} \cdot 8$

- entropy of … which takes values in the alphabet A and is distributed according to $P_X : A \to [0,1]$ is defined by:

$$H(X) := \sum_{x \in A} P_X(x) \log \frac{1}{P_X(x)}$$

$$= \mathbb{E}\Big[\underbrace{\log \frac{1}{P_X(X)}}_{\text{"surprise" - is larger when } P_X(x) \text{ is smaller}}\Big]$$

- properties:
  1) non-negative:
     $P_X(x) \in [0,1]$
     $[P_X(x)]^{-1} \geq 1$
     $\Rightarrow \log \frac{1}{P_X(x)} \geq 0$
  2) concave in $P_X(x)$
  3) $H(X) \leq \log_2 |A|$
  4) For a fixed A, entropy is maximized by a uniform distribution over A.

Joint Entropy

$$H(X,Y) := \sum_{x,y} P_{X,Y}(x,y) \log \frac{1}{P_{X,Y}(x,y)}$$

$$= \mathbb{E}\left[\log \frac{1}{P_{X,Y}(X,Y)}\right]$$

Conditional Entropy

$$H(Y|X) := \sum_{x \in \mathcal{X}} P_X(x) \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log \frac{1}{P_{Y|X}(y|x)}$$

$$= \sum_{x \in \mathcal{X}} P_X(x) H(Y|X=x)$$

$$= \mathbb{E}\left[\log \frac{1}{P_{Y|X}(Y|X)}\right]$$

Properties of Joint & Cond. Entropy

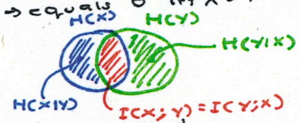
- $H(Y|X) = H(X,Y) - H(X)$
- chain rule of entropy:
  $H(X,Y) = H(X) + H(Y|X)$
- $H(Y|X) \leq H(Y)$
- $H(X,Y) \leq H(X) + H(Y)$

Mutual Information

- if X & Y aren't independent, they provide some information about each other, which mutual information measures:

$I(X;Y) := H(X) - H(X|Y)$
$= H(X) + H(Y) - H(X,Y)$
$= H(Y) - H(Y|X)$

↳ mutual information is symmetric
$I(X;Y) = I(Y;X)$
→ always non-negative
→ equals 0 iff X & Y independent



Baye's rule for Conditional Entropy

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

↳ if Y conditionally independent of Z given X, we have:
$H(Y|X,Z) = H(Y|X)$

↳ also:
$H(X,E|Y) = H(X|Y) + H(E|X,Y)$

- source coding: mapping from symbols x in the source alphabet $\mathcal{X}$ to bit strings
- $\ell(x) :=$ length of binary string description for X
- for a sequence of symbols, the binary description of this sequence has length $\ell(X_1, X_2, X_3, \ldots, X_n)$ & the average bits per symbol's $\frac{\ell(X_1, X_2, \cdots, X_n)}{n}$.

- Source coding Thm: For iid $X_1, \ldots, X_n$ and arbitrarily small $\varepsilon > 0$, there's a source coding scheme for which

$$\lim_{n \to \infty} \mathbb{E}\left[\tfrac{1}{n} \ell(X_1, \ldots, X_n)\right] \leq H(X) + \varepsilon$$

conversely, there's no source coding that can achieve less than $H(X)$ bits per symbol in expectation

Typical Sets & Asymptotic Equipartition Property (AEP)

- typical set: as n grows large, all the probability concentrates in an exponentially smaller subset of sequences in $\mathcal{X}^n$; this is the typical set:
  $\varepsilon > 0$, $n \geq 1$

$$A_\varepsilon^{(n)} := \{(x_1, \ldots, x_n) : P(x_1 \cdots x_n) \geq 2^{-n(H(X)+\varepsilon)}\} \subset \mathcal{X}$$

- Asymptotic Equipartition Property (AEP):
  if $(X_n)_{n \geq 1} \overset{iid}{\sim} P_X$, then:
  $-\frac{1}{n} \log P(X_1, \ldots, X_n) \to H(X)$ in prob
  ↳ ie, intuitively:
  $P(X_1, \ldots, X_n) \approx 2^{-nH(X)}$
- properties of typical set:
  1) $P\{(X_1 \cdots X_n) \in A_\varepsilon^{(n)}\} \to 1$ as $n \to \infty$
  2) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$

Communication:

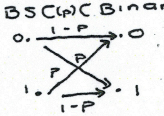
- Shannon's block diagram of comm. system
  ↳ n := "block length" = # of channel uses
  ↳ R := rate = (# info bits)/(# channel uses) = $nR/n = R$

$M \sim \text{Unif}(\{1, \ldots, 2^{nR}\})$ → encoder → $X^n(M)$
message = nR bits (random string)
codeword = n channel inputs = encoded msg

→ Noisy channel $P(Y|X)$ → $Y^n = (Y_1, \ldots, Y_n)$ → decoder → $\hat{M}(Y)$
channel output = received sequence

- $P_e^{(n)} := \mathbb{P}\{M \neq \hat{M}(Y^n)\}$ = probability receiver makes a mistake

BSC(p) (Binary Symmetric Channel):

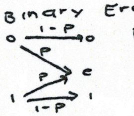


$$P_{Y|X}(y|x) = \begin{cases} (1-p) & \text{if } y = x \\ p & \text{if } y \neq x \end{cases}$$

$I(X;Y) = H(Y) - H(Y|X)$
$= H(Y) - \sum p(x) H(Y|X=x)$
$= H(Y) - \sum p(x) H(p)$
$= H(Y) - H(p)$
$\leq 1 - H(p)$ → achieved when input distribution is uniform

$\Rightarrow C = 1 - H(p) = 1 - [-p\log p - (1-p)\log(1-p)]$
$= 1 + p\log p + (1-p)\log(1-p)$

Binary Erasure Channel (BEC(p))



$$P_{Y|X}(y|x) = \begin{cases} (1-p) & \text{if } y = x \\ p & \text{if } y = e \end{cases}$$

$C = 1 - p$

DTMC (Discrete Time MC)

- at each time step n, the MC has a state X_n, belonging to a finite set S of possible states
- described in terms of transition probabilities P_{ij} (probability of transitioning from state i to state j)
  $P_{ij} = P(X_{n+1} = j \mid X_n = i) \quad i,j \in S$
- Markov property (memorylessness):
  $P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i) = P_{ij}$

$\sum_{j=1} P_{ij} = 1 \quad \forall i$ } "row stochastic" → rows $\sum$ to 1

- Transition probability matrix: (i^{th} row, j^{th} col = P_{ij})

$$\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ P_{21} & & & \\ \vdots & \vdots & \ddots & \\ P_{m1} & P_{m2} & & P_{mm} \end{bmatrix}$$

- n-step transition probabilities:
  $r_{ij}(n) = P[X_n = j \mid X_0 = i]$
  ↳ ie, the probability that after n time-steps, the state will be j, given that the current state is i
- Chapman-Kolmogorov Equation for the n-step transition probabilities:

$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1) P_{kj} \quad \leftarrow n > 1, \forall i,j$

$r_{ij}(1) = P_{ij}$ (this is a recursive formula)

- k-step transition matrix:
  $P_k(x,y) := P(X_k = y \mid X_0 = x)$
  $= \sum_{x_1, \ldots, x_{k-1}} \mathbb{P}[X_k = y, X_{k-1} = x_{k-1}, \ldots, X_1 = x_1 \mid X_0 = x]$
  $= \sum P(x, x_1) P(x_1, x_2) \cdots P(x_{k-1}, y) = P^k(x,y)$

- channel capacity (C): maximum rate that can be transmitted with perfect transmission

$$C = \max_{P_X(x)} I(X;Y)$$

- Channel coding Theorem: Any rate below the channel capacity C is achievable. Conversely, any sequence of codes with $P_e^{(n)} \to 0$ as $n \to \infty$ has a rate $R \leq C$

Chapman-Kolmogorov:

$P_{k+\ell} = P_k P_\ell \quad \forall k, \ell \in \mathbb{N}$

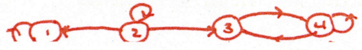
- denote distribution of X_n by row vector π_n & the distribution of
  $\pi_n = \pi_0 P^n$
- stationary distribution: $\pi_0 = \pi_0 P$
  $\Rightarrow \pi_n = \pi_0$

→ $\pi P = \pi$ more explicitly is:
$\pi(x) = \sum_{y \in \mathcal{X}} \pi(y) P(y,x) \quad \forall x \in \mathcal{X}$

Balance Equations
↳ set of $|\mathcal{X}|$ eqns (linear)
↳ normalization condition:

$$\sum_{x \in \mathcal{X}} \pi(x) = 1$$

can be solved using GE



- class properties: a property shared by all states in a class
- Accessible $(i \to j)$: j is accessible from i if there's a path from i to j (ie, $P_{ij}^n > 0$) for some $n \geq 0$.
- i & j communicate $(i \leftrightarrow j)$ iff $i \to j$ & $j \to i$
- Class of states = equivalence class under "$\leftrightarrow$"
  ↳ "$\leftrightarrow$" partitions S into equivalence classes (classes of states)
- irreducible MC - only has 1 class
- class properties:

  each state visited finite # of times
  1) Transience: the probability that you will return to a state x after leaving that state < 1.
  2) Recurrence: probability of returning to a state x after leaving it $= 1$.
     define $T_i := \min\{n \geq 1 : X_n = i\}$ = first re-entry time into state i
     A) Positive recurrent:
        $\mathbb{E}[T_i \mid X_0 = i] < \infty$

     revisits a state ~~daily~~ often but takes ∞ time to return to any given state
     B) Null recurrent:
        $\mathbb{E}[T_i \mid X_0 = i] = +\infty$
  C) Periodicity: For $i \in S$
     $\text{period}(i) = \gcd\{n \geq 1 : P_{ii}^n > 0\}$
     i) Aperiodic: if any (& ∴ all) states in a class have period = 1

- Big THM for MC's (long-term behavior):
  - Let $(X_n)_{n \geq 0}$ be an irreducible MC. Exactly one of the following is true:
    1. All states are transient / null-recurrent. In this case, no stationary distribution exists and $\lim_{n \to \infty} P_{ij}^n = 0 \quad \forall j, i$
    2. All states are positive recurrent. In this case, stationary distribution exists & is unique and satisfies

$$\pi_j = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P_{ij}^k$$

$$= \frac{1}{\underbrace{\mathbb{E}[T_j \mid X_0 = j]}_{\text{expected amount of time to return to state } j \text{ given you start in state } j}} \quad \forall i,j \in S$$

↳ Moreover, if the chain is aperiodic, then $P_{ij}^n \to \pi_j$ as $n \to \infty \quad \forall i,j \in S$

* Key IDEA for BIG THM: try to compute the stationary distribution to see if you're in 1 or 2.

- Balance Equations in general:
  ↳ used to compute ~~SD~~

$$\underbrace{\pi_j}_{\text{flow out of state } j} = \underbrace{\sum_i \pi_i P_{ij}}_{\text{flow into state } j} \quad \forall j$$

- Detailed Balance Equations
  - used for reversible, irreducible MCs.
  $\pi_i P_{ij} = \pi_j P_{ji}$

Probability

Chain rule:

$P(A,B) = P(A|B)P(B) = P(B|A)P(A)$

$P(A_1, A_2, \dots, A_k) = P(A_1)P(A_2|A_1) \cdots P(A_k|A_1, \dots A_{k-1})$

Conditional Prob/ Bayes:

$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A,B)}{P(B)}$

Inference by Enumeration

- 3 types of variables:
  1. Query variables (Q_i) unknown, appear on left side of the conditional in desired prob. dist.
  2. Evidence variables (e_i): observed, known values. Appear on right side of desired prob. dist
  3. Hidden variables: present in joint distribution but not the desired prob. dist.
- Algorithm:
  1. Collect all rows consistent w/observed evidence vars.
  2. sum out/marginalize all hidden variables
  3. Normalize

Mutual Independence $A \perp\!\!\!\perp B$

$A \perp\!\!\!\perp B \Rightarrow P(A,B) = P(A)P(B)$

$\Rightarrow P(A|B) = P(A)$

$\Rightarrow P(B|A) = P(B)$

Conditional Independence

$A \perp\!\!\!\perp B | C \equiv B \perp\!\!\!\perp A | C$

$\Rightarrow P(AB|C) = P(A|C)P(B|C)$

$\Rightarrow P(B|A,C) = P(B|C)$

Utility: Fcn from outcomes (states of world) to real #s that describe agent preferences

- Principle of maximum utility: rational agents must always select the action that max'es their expected utility
- lottery L has diff prob assoc. w/diff prizes:
  - ex: receive A wp p, B o/w: $L = [p, A; (1-p), B]$
- A prefered to B: $A \succ B$
- A & B indifferent: $A \sim B$

Axioms of Rationality:

- orderability: $(A \succ B) \vee (B \succ A) \vee (A \sim B)$
- transitivity: $(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$
- continuity: $(A \succ B \succ C) \Rightarrow \exists p\, [p, A; 1-p, C] \sim B$
- substitutibility: $(A \sim B) \Rightarrow [p, A; (1-p), C] \sim [p, B; (1-p), C]$
- Monotonicity: $(A \succ B) \Rightarrow (p \geq q) \Leftrightarrow [p, A; 1-p, B] \succeq [q, A; (1-q), B]$

Bayes Nets

- n variable, d values (domain size=d), k parents (max) — BN size: $O(nd^k)$
- joint distribution table: d^n entries for inference by enumeration → avoid this by using Bayesian inference
- def: a Bayes net consists of a
  - DAG (directed acyclic graph) of nodes; one per variable X
  - conditional distribution for each node $P(X | A_1 \dots A_n)$, where A_i is the i^{th} parent of X, stored as a CPT. → each CPT has n+2 columns
- prob rule for ~~___~~ Bayes Nets:

$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P(X_i | parents(X_i))$

- 2 rules for BNs:
  1. Each node conditionally indep. of all its ancestor nodes (non-descendants) in the graph, given all of its parents
  2. Each node is conditionally indep. of all other variables given its Markov blanket. (Mk blanket = parents, children, & children's other parents)

D-seperation

- casual chains $X \to Y \to Z$
  - expresses the joint distribution $P(x,y,z) = P(z|y)P(y|x)P(x)$
  - $\to P(X|Z,Y) = P(X|Y) \Rightarrow X \perp\!\!\!\perp Z | Y$
- common cause $X \leftarrow Y \to Z$
  - $P(x,y,z) = P(x|y)P(z|y)P(y)$
  - $X \perp\!\!\!\perp Z | Y$
- common effect $X \to Y \leftarrow Z$
  - $P(x,y,z) = P(y|x,z)P(x)P(z)$
  - $X \perp\!\!\!\perp Z$

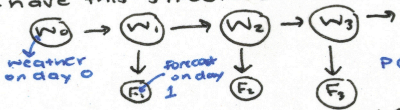
- analogous to chain-like, infinite Baye's net
- Mini fwd algorithm: by marginalization, we know $P(W_{i+1}) = \sum_{w_i} P(w_i, W_{i+1})$

Thus: $P(W_{i+1}) = \sum_{w_i} P(W_{i+1}|w_i)P(w_i)$

- joint distribution for a Markov model: $P(X_0, \dots, X_T) = P(X_0) \prod_{t=1:T} P(X_t | X_{t-1})$

Hidden Markov Models (HMMs)

- have this structure:



- joint distribution for HMM: $P(X_0, X_1, \dots, X_T, E_T) = P(X_0) \prod_{t=1:T} P(X_t|X_{t-1})P(E_t|X_t)$

W_i := state variable: rv encoding belief at a timestep

F_i := evidence variable: rv encoding observation at a timestep

$P(W_0)$ = initial distribution

$P(F_i | W_i)$ = sensor model

$P(W_{i+1} | W_i)$ = transition model

- good to use HMMs when evidence var at a timestep indep. of everything else given that hidden var at that timestep, & when the hidden var is time-varying

Forward Algorithm:

- belief distribution at time i given $f_1, \dots, f_i$

$B(W_i) = P(W_i | f_1, \dots, f_i)$

$B'(W_i) = P(W_i | f_1, \dots, f_{i-1})$

$= \sum_{w_i} P(W_i | w_{i-1}) B(w_{i-1})$

- Forward Algorithm:

$B(W_{i+1}) \propto \underbrace{P(f_{i+1}|W_{i+1})}_{\text{observation update}} \underbrace{\sum_{w_i} P(W_{i+1}|w_i)B(w_i)}_{B'(W_{i+1})}$ ← time elapse update

$B'(X_t) = P(X_t | X_{t-1}=0) B(X_{t-1}=0) + P(X_t | X_{t-1}=1) B(X_{t-1}=1)$

$B(X_1) = \frac{P(E_1 = e | X_1) B'(X_1)}{\sum_{x_1} P(E_1 = e | X_1 = x_1) B'(X_1 = x_1)}$

Stationary Distribution

- satisfies $P_\infty(X_t) = \sum_{x_t} P(X_{t+1} | X_t = x_t) P_\infty(X_t = x_t)$

Inference Tasks

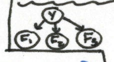
- Filtering: $P(X_t | e_{1:t})$
  - HMM analog to BN sampling
  - n particles, d posns

Other repr. for fwd algo: (to approx. $P(X_t | e_1, \dots, e_t)$)

1. Elapse time: $P(X_t | e_{1 \dots t-1}) = \sum_{x_{t-1}} P(X_t | x_{t-1}) P(x_{t-1} | e_{1 \dots t-1})$
2. Observe: $P(X_t | e_{1 \dots t}) = \frac{P(e_t | X_t) P(X_t | e_{1 \dots t-1})}{\sum_{x_t} P(e_t | x_t) P(x_t | e_{1 \dots t-1})}$

- supervised learning: infer relationship b/w input data & output data, so we can predict outputs for unseen inputs in the future. used for classification problems, where we have labels
- unsupervised learning: don't have corresponding output data
- split data into training & validation, then make predictions on the test set

Naive Bayes



prediction($f_1, \dots, f_n$):

$= \arg\max_y P(Y=y | F_1 = f_1, \dots, F_N = f_n)$

$= \arg\max_y P(Y=y, F_1 = f_1 \dots F_N = f_n)$

$= \arg\max_y P(Y=y) \prod_{i=1}^{n} P(F_i = f_i | Y=y)$

$P(Y, F_1 = f_1, \dots, F_N = f_n) = \begin{bmatrix} P(Y=y_1, F_1 = f_1, \dots F_n = f_n) \\ \vdots \\ P(Y=y_k, F_1 = f_1, \dots F_n = f_n) \end{bmatrix}$

$= \begin{bmatrix} P(Y=y_1) \prod P(F_i = f_i | Y = y_1) \\ \vdots \\ P(Y=y_k) \prod P(F_i = f_i | Y = y_k) \end{bmatrix}$

Parameter estimation

- N sample pts/observations $x_1, \dots, x_N$ & believe that data was drawn from distribution parameterized by unknown value θ.
  - MLE can estimate this θ

Maximum Likelihood Est (MLE)

- makes the assumptions:
  - each x_i identically dist.
  - each x_i cond. indep. of the others
  - all values of θ are equally likely before we've seen any data (uniform prior)

likelihood $\mathcal{L}(\theta)$

$\mathcal{L}(\theta) = P_\theta(x_1, \dots, x_N) = \prod_{i=1}^{N} P_\theta(x_i)$

- MLE for θ satisfies $\frac{\partial}{\partial \theta} \mathcal{L}(\theta) = 0$
- given n observations, $P_{MLE}(x) = \frac{count(x)}{\#samples}$

Laplace smoothing:

- way to counteract overfitting by assuming you've seen k more instances of each outcome:

$P_{LAP,k}(x) = \frac{count(x) + \overbrace{k}^{\text{"strength of prior"}}}{\underbrace{N}_{\#samples} + k \underbrace{|X|}_{\#classes}}$

$P_{LAP,k}(x|y) = \frac{count(x,y) + k}{count(y) + k|X|}$

$P_{LAP,\infty}(x) = \frac{1}{|X|}$

Linear Regression

- aka least squares
- regression probs have output that's a continuous variable
- not used for classification

- $\vec{x} \in \mathbb{R}^n$:= set of n features
- $h_w(\vec{x}) = w_0 + w_1 x_1 + \dots + w_n x_n$ (linear model to pred. the output; w_i are weights we want to estimate)
- $loss(h_w) = \frac{1}{2} \| \vec{y} - X\vec{w} \|_2^2 = \frac{1}{2} \sum_{j=1}^{n} (y_j - h_w(\vec{x}_j)^2)$
- $\hat{w} = (X^T X)^{-1} X^T \vec{y}$ (best weight estimate)
- $h_{\hat{w}}(\vec{x}) = \hat{w}^T \vec{x}$ (prediction on new data)

Linear Classifiers: do classification using linear combo of features (aka activation).

$activation_w(\vec{x}) = h_w(\vec{x}) = \sum_i w_i f_i(\vec{x}) = \vec{w}^T \vec{f}(x)$

Binary:

$classify(\vec{x}) = \begin{cases} + & \text{if } h_w(\vec{x}) > 0 \\ - & o/w \end{cases}$

Sampling

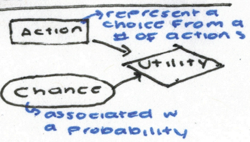
1. Prior sampling P: generate complete samples from $P(x_1, x_2, \dots, x_n)$
   - for $i = 1, 2, \dots, n$ (in topological order): sample x_i from $P(X_i | parents(X_i))$; Return $(x_1, \dots, x_n)$
2. Rejection sampling $P(Q|e)$: reject samples that don't match e
   - generate samples, discard samples inconsistent w/evidence. calculate prob.
3. Likelihood weighting $P(Q|e)$: weight samples by how much they predict e
   - input evidence $e_1, \dots, e_k$
   - for $i = 1, 2, \dots, n$:
     - sample x_i from $P(X_i | parents(X_i))$
     - if x_i inconsistent w/evidence reject & return: no sample generated from this cycle
   - return $x_1, \dots, x_n$
   - ex: if we want $P(C | r, w)$, we can discard samples with ¬r or ¬w
4. Gibbs sampling $P(Q|e)$: wander around in e space & ~~wt~~ avg what you see
   - input evidence $e_1, \dots, e_k$
   - $w = 1.0$
   - for $i = 1, \dots, n$
     - if X_i an evidence variable
       - x_i = observed value$_i$ for X_i
       - $w = w * P(x_i | Parents(X_i))$
     - else:
       - sample x_i from $P(X_i | parents(X_i))$
   - return $x_1, x_2, \dots, x_n, w$
   - note: downstream vars influenced by upstream but not vice versa.
   - note: all samples used
   - states = complete assgn to all vars
   - fix e_i vars; other vars can change
   - gen subsequent states by picking non-evidence vars & sampling for it, cond on other vars
   - $P(X_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$: $P(X_i | \text{markov blanket}(X_i))$
   - considers down & upstream evidence

Utility (con't)

- if all 5 axioms of rationality are satisfied by an agent, then it's guaranteed that the agent ~~___~~ maximizes utility,
  - ⇒ ∃ a real-valued utility fcn U s.t.:
  - $U(A) \geq U(B) \Leftrightarrow A \succeq B$
  - $U([p_1, S_1; \dots; p_n, S_n]) = \sum_i p_i U(S_i)$
- risk-neutral: indiff btwn receiving flat payment & participating in lottery
- risk-averse: prefers flat payment
- risk-~~seeking~~: prefers lottery
- Given lottery $L = [p, \$X; (1-p), \$Y]$:
  - expected monetary value $EMV(L) = pX + (1-p)Y$
  - utility $U[L] = pU(\$X) + (1-p)U(\$Y)$
    - usually $U(L) < U(EMV(L))$ ⇒ ppl are risk-averse
  - certainty equivalent: $CE(L) \sim L$
  - insurance premium: $EMV(L) - CE(L)$
    - if ppl were risk-neutral, this would = 0

- if we assume stationary preferences: $[a_1, a_2, \dots] \succ [b_1, b_2, \dots] \Leftrightarrow [c, a_1, a_2, \dots] \succ [c, b_1, b_2, \dots]$ then there's only 1 way to define utilities:
  - Additive Discounted Utility: $U([r_0, r_1, r_2, \dots]) = r_0 + \gamma r_1 + \gamma^2 r_2 + \dots$ where $\gamma \in [0,1]$ is the discount factor

Desicion Networks



Expected Utility $EU(a|e) = \sum_{x_1, \dots, x_n} P(x_1, \dots, x_n | e) U(a, x_1, \dots, x_n)$

"expected utility of action a given evidence e & n chance nodes"

Maximum Expected Utility $MEU(E=e) = \max_a EU(A=a | E=e)$

- choose the action that maximizes the EU

MEU if observing evidence e': $MEU(e, e') = \max_a \sum_s P(s|e,e') U(s,a)$

Value of Perfect Information: $VPI(E_i | e) = [\sum_{e_i} P(e_i|e) \max_a EU(a|e_i, e)] - \max_a EU(a|e)$

Value of Perfect Information (VPI) Properties

1. Non-negativity: observing new evidence can only help you/be irrelevant: $\forall E_i, e \; VPI(E_i | e) \geq 0$
2. Non-additivity: observing evidence E_j might affect how much we care abt E_k. $VPI(E_j, E_k | e) \neq VPI(E_j | e) + VPI(E_k | e)$
3. order independence: order of observing evidence doesn't matter

Concentration Inequalities

- **Markov's inequality:**
  - if $X \geq 0$ is a r.v. then
    $P\{X \geq \lambda\} \leq \frac{\mathbb{E}[X]}{\lambda} \quad \lambda \geq 0$
  - can take fcn of X instead:
    $P[|X-\mathbb{E}[X]| > \lambda] \leq \frac{\mathbb{E}[|X-\mathbb{E}[X]|]^k}{\lambda^k}$
- **Chebyshev's inequality:**
  $P[|X-\mathbb{E}[X]| \geq \lambda] \leq \frac{Var(X)}{\lambda^2}$
  - can use MGF to take in more info:
    $P[X \geq \lambda] = P[e^{tX} \geq e^{t\lambda}]$
    $\underset{\text{Markov}}{\leq} \frac{\mathbb{E}[e^{tX}]}{e^{t\lambda}} = \frac{M_x(t)}{e^{t\lambda}}$
- **Chernoff Bound:**
  $P[X \geq \lambda] \leq \min_{t \geq 0} \frac{M_x(t)}{e^{\lambda t}}$

Weak Law of Large Numbers:
$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > \varepsilon\right\} = 0$

Properties of Probability Measures:

1. $P(A^c) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. $A \subseteq B \Rightarrow P(A) \leq P(B)$
4. $P(A) \leq 1$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Expected Sorting Distance

$\mathbb{E}[\sum_{i=1}^{n} |a_i - i|]$

$= \sum_{i=1}^{n} \mathbb{E}[|a_i - i|]$

$\mathbb{E}[|a_i - i|] = \sum_{k=1}^{n} \frac{1}{n}|k-i|$

$= \frac{1}{n}\sum_{k=1}^{n-i} k + \frac{1}{n}\sum_{k=1}^{i-1} k$

$= \frac{(n-i)(n-i+1) + (i-1)i}{2n}$

$\mathbb{E}[\sum_{i=1}^{n} |a_i - i|] = \frac{n^2-1}{3}$

Joint Density for Exp Distr.

$X \sim Exp(\lambda) \quad Y \sim Exp(\mu)$ indep.

$\mathbb{P}[X < Y] = \int_0^\infty \underbrace{P[X<y \mid Y=y]}_{P[X<y] \text{ bc of indep.}} \cdot f_Y(y)\,dy$

$= \int_0^\infty (1 - e^{-\lambda y})(\mu e^{-\mu y})\,dy$

$= \frac{\lambda}{\lambda+\mu}$ $\quad \mathbb{E}[Z := \min(X,Y)] = \frac{1}{\lambda+\mu}$, $Z \sim Exp(\lambda+\mu)$

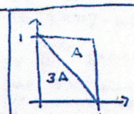
$X_1, \ldots, X_n$ indep & exponentially distr. w/params $\lambda_1, \ldots, \lambda_n$

$X := \min_{1\leq k\leq n} X_k \sim Exp(\sum_{j=1}^{n} \lambda_j)$

$P[X \geq x] = P[X_1 \geq x, \ldots, X_n \geq x]$
$= \prod_{k=1}^{n} P[X_k \geq x] = \prod_{k=1}^{n} e^{-\lambda_k x}$
$= e^{-x\sum_{k=1}^{n}\lambda_k}$

$P[X_i = \min_{1\leq k\leq n} X_k] = \frac{\lambda_i}{\sum_{j=1}^{n}\lambda_j}$

$X = (X_1, \ldots, X_n)$
$X_1, \ldots, X_n$ iff $\forall t \in \mathbb{R}^n$:
$M_X(t) = \prod_{i=1}^{n} M_{X_i}(t_i)$



1. X, Y indep? No. $X=0, Y=\frac{1}{2}$ but $X=\frac{1}{2}, Y<\frac{1}{2}$
2. $A = ?$
   $\frac{1}{2}(3A) + \frac{1}{2}(A) = 1$
   $A = \frac{1}{2}$
3. $f_X(x) = ?$
   $f_X(x) = \int_0^1 f_{X,Y}(x,y)\,dy$
   $= \int_0^{1-x} \frac{3}{2}\,dy + \int_{1-x}^{1} \frac{1}{2}\,dy$
   $= \frac{3}{2}(1-x) + \frac{1}{2}x = \frac{3}{2} - x$
4. $\mathbb{E}[Y \mid X=x] = ?$
   $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
   $\mathbb{E}[Y \mid x] = \int_0^{1-x} y \cdot \frac{3A}{3/2 - x}\,dy + \int_{1-x}^{1} y\frac{A}{3/2-x}\,dy$
   $= \frac{3A(1-x)^2}{3-2x} + \frac{A(1-(1-x)^2)}{3-2x} = \frac{3-4x+2x^2}{2(3-2x)}$
5. $\mathbb{E}[X-Y \mid X+Y] = ?$
   $\mathbb{E}[X \mid X+Y] = \frac{X+Y}{2} = \mathbb{E}[Y \mid X+Y]$
   $\mathbb{E}[X-Y \mid X+Y] = 0$

Poisson Properties

X, Y indep $X \sim Pois(\lambda)$, $Y \sim Pois(\mu)$

→ show the distribution of $(X+Y)$ is Poisson:

$P[X+Y=z] = \sum_{x=0}^{z} P[X=x]P[Y=z-x]$

$= \sum_{x=0}^{z} \frac{\lambda^x}{x!} e^{-\lambda} \cdot \frac{\mu^{z-x} e^{-\mu}}{(z-x)!}$

$= \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x} \frac{e^{-(\lambda+\mu)}}{z!}$

$= \frac{e^{-(\lambda+\mu)}}{z!}(\lambda+\mu)^z$

$\Rightarrow X+Y \sim Pois(\lambda+\mu)$

$P[X=x \mid X+Y=z] = ?$

$= \frac{P(X=x, X+Y=z)}{P[X+Y=z]}$

$= \frac{P[X=x]P[Y=z-x]}{\left[\frac{e^{-(\lambda+\mu)}}{z!}\cdot(\lambda+\mu)^z\right]}$

$= \binom{z}{x}\left(\frac{\lambda}{\lambda+\mu}\right)^x\left(\frac{\mu}{\lambda+\mu}\right)^{z-x}$

$\Rightarrow X \mid X+Y=z \sim Bin(z, \frac{\lambda}{\lambda+\mu})$

Coupon Collector

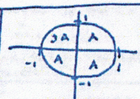
- n diff types of coupons, each box contains 1 coupon.
  $X :=$ # boxes bought until one of each obtained:
  $\mathbb{E}[X] = n H_n = n(\sum_{i=1}^{n} \frac{1}{i})$

Café 126

$N \sim Pois(\mu) \quad X_i \sim Exp(\lambda)$
$Y = \sum_{i=1}^{N} X_i$

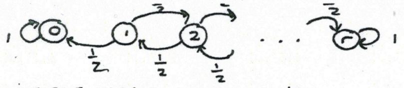
$\mathbb{E}[Y] = \mathbb{E}[\sum_{i=1}^{N} X_i] = \mathbb{E}[\mathbb{E}[\sum_{i=1}^{N} X_i \mid N]]$
$= \mathbb{E}[N \cdot \frac{1}{\lambda}] = \frac{\mu}{\lambda}$

$Var(Y) = \mathbb{E}[var(Y|N)] + var[\mathbb{E}[Y|N]]$
$= \mathbb{E}[N]\,var(X_1) + var(N\,\mathbb{E}[X_1])$
$= \mathbb{E}[N]\,var(X_1) + \mathbb{E}[X_1]^2\,var(N)$
$= \frac{\mu}{\lambda^2} + \frac{\mu}{\lambda^2} = \frac{2\mu}{\lambda^2}$



$x^2 + y^2 = 1$

1. $A = ?$
   $A(\frac{\pi}{4}) + 3A(\frac{\pi}{4}) + A(\frac{\pi}{4}) + 3A(\frac{\pi}{4}) = 2\pi A = 1$
   $A = \frac{1}{2\pi}$
2. $f_x = ?\ f_y = ?$
   $f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)\,dy = \int_{-\sqrt{1-x^2}}^{0} 3A\,dy + \int_0^{\sqrt{1-x^2}} A\,dy = \frac{2}{\pi}\sqrt{1-x^2}$
   $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$ by rotational symmetry
3. X, Y indep?
   X & Y indep iff $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \quad \forall (x,y)$
   → consider x, y in 1st quad.
   $f_X(x)f_Y(y) = f_X(x)f_Y(-y) = 3A$
   but $f_{x,y} = A \Rightarrow$ dependent.



Time until game ends

$A = \{0, R\}$

$t(i) = \begin{cases} 0 & i \in \{0, R\} \\ 1 + \frac{1}{2}(t(i-1) + t(i+1)) & i \in \{1, \ldots, R-1\} \end{cases}$

$t(k) = k(r-k)$

Hitting Prob

$h(a) = 1 \quad h(b) = 0$

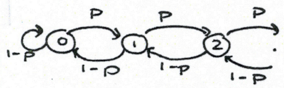
$h(i) = \sum_{j\in S} h(j)P_{ij} \quad i \notin \{a, b\}$

$\Rightarrow a = R \quad b = 0$

$h(i) = \frac{1}{2}[h(i-1) + h(i+1)] \quad i \in \{1, \ldots, R-1\}$

$h(k) = \frac{k}{r} \quad 0 \leq k \leq r$

Birth-Death Process



if
- $p = \frac{1}{2}$ MC is null-recurrent
- $p < \frac{1}{2}$ MC is positive-recurrent
  ↳ fwd rate less than backwards rate
- $p > \frac{1}{2}$ MC is transient

Two DTMCs associated with CTMC

1) Jump chain
   - ↳ normalize outgoing edges for each node
   - ↳ no self-loops
   - ↳ equivalent to CTMC w/o exponential clk
2) Chain with same SD
   - ↳ normalize all edges by the same constant where constant = max outflow rate of any node
   - ↳ if the outflow at a node doesn't Σ to 1, the remaining amt goes into a self-loop

CTMCs

- $\pi Q = 0$.
- $\alpha(s) = \underbrace{\frac{1}{\lambda_s}}_{\text{bc our expected time is } \frac{1}{\lambda_s}} + \sum_{j\in S} \frac{Q_{sj}}{\lambda_s}\alpha(j)$

  $Exp\,\alpha(\lambda_s)$

M/M/S Queue

$\pi(1) = \frac{\lambda}{\mu}\pi(0) \qquad \pi(s+i) = \frac{\lambda}{s\mu}\pi(i)$

$\pi(i) = \left(\frac{\lambda}{i\mu}\right)^{i-1}\pi(0) \qquad \Sigma\ ?$

$\sum_{i=0}^{\infty} \pi(i) = \pi(0) + \frac{\lambda}{\mu}\pi(0) + \sum_{i=2}^{\infty}\left(\frac{\lambda}{i\mu}\right)^{i-1}\pi(0) = 1$

Probability Axioms

1) $P(A) \geq 0 \quad \forall A$
2) A_i's disjoint: $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
3) $P(\Omega) = 1$

Discrete Prob. Law

For finite Ω, the P of some event $\{s_1, s_2, \dots, s_n\} = P\{s_1\} + P\{s_2\} + \dots + P\{s_n\}$

Discrete Unif P Law

$P(A) = \dfrac{\#\text{ elems in } A}{n}$

Properties of Prob Laws

For events A, B, C ($A, B, C \in F$)

1) $A \subset B \Rightarrow P(A) \leq P(B)$
2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3) $P(A \cup B) \leq P(A) + P(B)$
4) $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$

Conditional Prob

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$P(A \cap B) = P(A|B)P(B)$

CP Mult Rule

$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots = P(A_n | \bigcap_{i=1}^{n-1} A_i)$

Total Probability Thm

$P(B) = P(A_1 \cap B) + \dots + P(A_n \cap B) = P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n)$

Bayes Rule

$P(A_i|B) = \dfrac{P(B|A_i)P(A_i)}{P(B)} = \dfrac{P(B|A_i)P(A_i)}{P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n)}$

Independence

$P(A|B) = P(A)$

$P(A \cap B) = P(A)P(B)$

Conditional Independence

$P(A \cap B | C) = P(A|C)P(B|C)$

$= \dfrac{P(A \cap B \cap C)}{P(C)}$

$= \dfrac{P(C)P(B|C)P(A|B \cap C)}{P(C)}$

$= P(B|C)P(A|B \cap C)$

$P(A|B \cap C) = P(A|C)$

↳ if $P(B \cap C) > 0$

For several events, independence is:

$P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$

Counting

- Principle: if there's a sequence of independent events, that can occur $a_1, a_2, \dots, a_n$ ways, the # of ways for all events to occur is $\prod_{i=1}^{n} a_i$.
- # of subsets of n element set: 2^n.

Permutations

- n distinct objects, $k \leq n$
- # ways we can pick k out of the n (ie, # permutations): $\dfrac{n!}{(n-k)!}$

Combinations (no ordering!)

- # k-element subsets of a given n-element set: $\binom{n}{k} = \dfrac{n!}{(n-k)!\,k!}$

Partitions

- n objects, $\sum_i n_i = n$, r disjoint groups, with i^{th} group containing n_i items:

$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\dots-n_{r-1}}{n_r} = \dfrac{n!}{n_1!\,n_2!\cdots n_r!} = \binom{n}{n_1, n_2, \dots, n_r}$

Random Variables

- Random Variable: real-valued function of the experimental outcome
- Discrete RVs: takes finite/countably infinite # of values, has a PMF.

Probability Mass Fcns

- gives the probability of each numerical value that the RV can take
- if x is any possible value of X, the probability mass of x is $p_X(x) = P(\{X = x\})$
- $\sum_x p_X(x) = 1$
- $P(X \in S) = \sum_{x \in S} p_X(x)$

Bernoulli Random Var

$X = \begin{cases} 1 & \text{if H} \\ 0 & \text{if T} \end{cases}$

$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ (1-p) & x = 0 \end{cases}$

$E[X] = p$

$var(X) = p(1-p)$

Binomial Random Var

- biased coin tossed n times, indep of other tosses.
- X = # Heads in n tosses
- $X \sim Bin(n, p)$

$p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

$E[X] = np \quad Var(X) = np(1-p)$

Geometric Random Variable

- repeatedly toss biased coin
- X = # tosses for 1^{st} H

$p_X(k) = (1-p)^{k-1} p$

$E[X] = \dfrac{1}{p} \quad Var(X) = (1-p)/p^2$

Poisson Random Variable

- probability of a given # of events (k) occurring in a fixed interval of time or space, if the events occur with a known constant mean rate ($E[X] = \lambda$) & indep. of the time since the last event.

$X \sim Pois(\lambda) \quad var(X) = \lambda$

$p_X(k) = \dfrac{\lambda^k e^{-\lambda}}{k!} \quad E[X] = \lambda$

Functions of RVs

$Y = g(X)$

- to calculate PMF of Y: $p_Y(y) = \sum_{\{x | g(x) = y\}} p_X(x)$

Expectation E (mean)

$E[X] = \sum_x x\, p_X(x)$

- n^{th} moment of an RV: $E[X^n] = \sum_x x^n p_X(x)$

Variance

$var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2 = \sum_x (x - E[X])^2 p_X(x)$

- always ≥ 0, measures dispersion of X around its mean

Standard Deviation σ

$\sigma_X = \sqrt{var(X)}$

LOTUS

$E[g(X)] = \sum_x g(x) p_X(x)$

↳ allows us to calc. the variance w/o using the PMF of $(X - E[X])^2$:

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

↳ also allows us to compute n^{th} moment:

$E[X^n] = \sum_x x^n p_X(x)$

Linear Fcns of RVs

$Y = \alpha X + \beta$

$E[Y] = \alpha E[X] + \beta$

$var(Y) = \alpha^2 Var(X)$

bc $var(Y) = \sum_x (\alpha x + \beta - E[\alpha X + \beta])^2 p_X(x)$

$= \alpha^2 \sum_x (x - E[X])^2 p_X(x)$

$= \alpha^2 Var(X)$

Joint PMFs

$p_{X,Y}(x,y) = P(\{X = x\} \cap \{Y = y\}) = P(X = x, Y = y)$

- determines the probability of any event that can be specified in terms of rv's X & Y. eg: $P((X,Y) \in A) = \sum_{(x,y) \in A} p_{X,Y}(x,y)$
- can use joint PMF to calculate marginal PMFs of X and Y: $p_X(x) = \sum_y p_{X,Y}(x,y)$

Discrete Uniform RV

- $X \sim Unif([a,b])$

$p_X(k) = \begin{cases} \dfrac{1}{b-a+1} & \text{if } k \in [a,b] \\ 0 & \text{o/w} \end{cases}$

$E[X] = (a+b)/2$

$E[X^2] = \dfrac{1}{n}\sum_{k=1}^{n} k^2 = \dfrac{1}{6}(n+1)(2n+1)$

$Var(X) = \dfrac{n^2 - 1}{12} = \dfrac{(b-a)(b-a+2)}{12}$

~~Expected~~ Functions of Multiple RVs

$E[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$

Conditional PMFs

- $p_{X|A} = P(X = x | A)$, $A \in F$

$= \dfrac{P(\{X = x\} \cap A)}{P(A)}$

- $P(A) = \sum_x P(\{X = x\} \cap A)$

$\sum_x p_{X|A}(x) = 1$

- X, Y RV's:

$p_{X|Y}(x|y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$ ← Joint PMF / Marginal PMF

- can use the conditional PMF to calc. joint PMF:

$p_{X,Y} = p_Y(y) p_{X|Y}(x|y) = p_X(x) p_{Y|X}(y|x)$

Conditional Expectation

- $E[X|A] = \sum_x x\, p_{X|A}(x|A)$
- $E[g(X)|A] = \sum_x g(x) p_{X|A}(x|A)$
- $E[X|Y = y] = \sum_x x\, p_{X|Y}(x|y)$
- ~~Total expectation thm:~~ ~~$E[X] = \sum_y p_Y(y) E[X|Y = y]$~~
- $A_1, \dots, A_n$ disjoint events, partitioning the sample space, then: $E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$

PMFs & Independence

- rV X indep. from event A if $P(X = x \cap A) = P(X = x)P(A) = p_X(x)P(A)$ or $p_{X|A}(x) = p_X(x) \quad \forall x$
- rv's X & Y independent if $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ (or equivalently): $p_{X|Y}(x,y) = p_X(x)$
- X & Y are conditionally independent if (for $A \in F$): $P(X = x, Y = y | A) = p_{X|A}(x) p_{Y|A}(y)$ ie $p_{X,Y|A}(x,y) = p_{X|A}(x) p_{Y|A}(y)$ or equivalently: $p_{X|Y,A} = p_{X|A}(x)$

Expectation of indep. rv's

$E[XY] = E[X]E[Y]$

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Variance of indep rv's

$Var(X+Y) = var(X) + var(Y)$

Continuous RVs

Probability density fcn:

$P(X \in B) = \int_B f_X(x)\,dx$

$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$

↳ $f_X(x)$ must be non-negative

$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

Continuous Uniform Random Variables

$f_X(x) = \begin{cases} c & 0 \leq x \leq 1 \\ 0 & \text{o/w} \end{cases}$

$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_a^b c\,dx = c$

$c = \dfrac{1}{b-a}$

$E[X] = \dfrac{a+b}{2}$

$Var(X) = \dfrac{(b-a)^2}{12}$

Expectation of Continuous RVs

- $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$
- $Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\,dx$
- if $Y = aX + b$: $E[Y] = aE[X] + b$, $Var(Y) = a^2 var(X)$

Normal/Gaussian RV

$f_X(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$

↳ $\sigma > 0$

$E[X] = \mu \quad var(X) = \sigma^2$

Standard Normal RV Φ

- normal RV w/ zero mean and unit variance
- CDF Φ

$\Phi(y) = P[Y \leq y] = P[Y < y] = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}}\,dt$

- to standardize Normal RV X, define $X \sim \mathcal{N}(\mu, \sigma^2)$

$Y = \dfrac{X - \mu}{\sigma} \quad var(Y) = 1 \quad E[Y] = 0$

- CDF calculation of Normal RV X w/mean μ and var σ^2:

$P[X \leq x] = P\left[\dfrac{X-\mu}{\sigma} \leq \dfrac{x-\mu}{\sigma}\right] = P\left[Y \leq \dfrac{x-\mu}{\sigma}\right] = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$

Cumulative Distr. Fcn (CDF)

$F_X(x) = P[X \leq x] = \begin{cases} \sum_{k \leq x} p_X(k) & \text{d…} \\ \int_{-\infty}^{x} f_X(t)\,dt & \text{cont…} \end{cases}$

- accumulates probability "up to" the value x

Properties:

- monotonically non-decreasing ↳ if $x \leq y \rightarrow F_X(x) \leq F_X(y)$
- $F_X(x) \rightarrow 0$ as $x \rightarrow -\infty$
- $F_X(x) \rightarrow 1$ as $x \rightarrow \infty$

Exponential RV

$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{o/w} \end{cases}$

$P[X \geq a] = e^{-\lambda a}$

$E[X] = 1/\lambda$

$Var(X) = 1/\lambda^2$

CDFs (cont)

$F_X(k) = \sum_{i=-\infty}^{k} p_X(i)$

$p_X(k) = P[X \leq k] - P[X \leq k-1] = F_X(k) - F_X(k-1)$

Conditional PDF & E given $A \in F$

- $P[X \in B | A] = \int_B f_{X|A}(x)\,dx$
- if A a subset of the real line w/ $P(X \in A) > 0$ then:

$f_{X|A}(x) = \begin{cases} \dfrac{f_X(x)}{P(X \in A)} & \text{if } x \in A \\ 0 & \text{o/w} \end{cases}$

and $P(X \in B | X \in A) = \int_B f_{X|A}(x)\,dx$

- $E[g(X)|A] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)\,dx$
- if $A_1, A_2, \dots, A_n$ disjoint w/ $P(A_i) > 0$ …:

$f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)$

$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$

$E[g(X)] = \sum_{i=1}^{n} P(A_i) E[g(X)|A_i]$