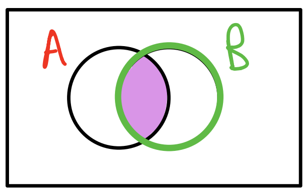
Final Review (post-midterm material)

Discrete Probability

- outcome, w
- Sample space, $\Omega = \{w_1, w_2, \dots\}$
- Event, $E \subseteq \Omega$
- Probability Space: $\Omega + \underbrace{P(w_i)}_{\text{probability}}$
- $$P[E] = \frac{\sum_{w \in E} P(w)}{\sum_{w \in \Omega} P(w)}$$
- Uniform Space: $P[E] = \frac{|E|}{|\Omega|}$ (counting problem!)

Conditional Probability

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$



Calculating probabilities:

Total Prob. Law: $P[E] = \sum P[\underbrace{E \cap C_i}_{\text{"pieces"}}] = \sum P[E \mid C_i]\, P[C_i]$

Bayes: $P[A \mid B] = \frac{P[B \mid A]\, P[A]}{P[B]}$

Independence:

$P[A \cap B] = P[A]\, P[B]$

$P[A \mid B] = P[A]$

$P[B \mid A] = P[B]$

Mutual $\Rightarrow$ Pairwise Independence

Random Variables (Discrete)

$$X = \begin{cases} x_1 & \text{w.p. } p_1 \\ x_2 & \text{w.p. } p_2 \\ \vdots \end{cases}$$

Notable Distributions:
- Bernoulli(p)
- Bin(n, p)
- Uniform$\{1, \dots, n\}$
- Geometric(p)
- Poisson(λ)
- Hypergeometric(N, B, n)

- PMF: $P[X=i]$
- $E[X] = \sum_i i \cdot P[X=i]$

Estimation

MMSE / Conditional Expectation: $E[Y \mid X]$

LLSE: $L[Y \mid X] = E[Y] + \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(X)}(X - E[X])$

Finding $E[X]$:

- Indicators + Linearity of Expectation

$$X = I_1 + I_2 + \dots + I_n$$

$$E[X] = E[I_1 + I_2 + \dots + I_n] = E[I_1] + \dots + E[I_n]$$

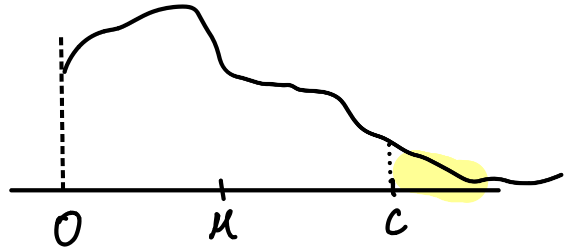
- Symmetry
- Smoothing Rule

$$E[E[X \mid Y]] = E[X]$$

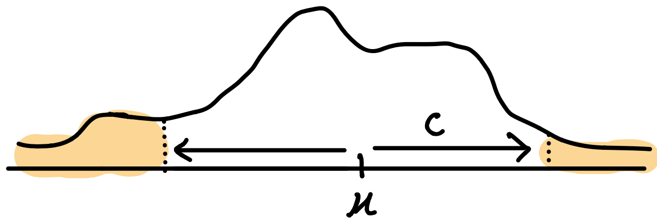
Variance: $\mathrm{Var}(X) = E[X^2] - (E[X])^2$

Concentration Bounds

- Markov



- Chebyshev



- Law of Large Numbers

Average of i.i.d. observations $\longrightarrow E[X_i]$

- Central Limit Theorem

Normalized Sum of i.i.d. observations $\longrightarrow \mathcal{N}(0,1)$

Random Variables (Continuous)

p.d.f., $f_X(x)$

c.d.f., $P[X \leq x] = F[x]$

$$f_X(x) = \frac{dF}{dx}$$

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

Notable Distributions
$$\begin{cases} \text{Uniform}[0, \ell] \\ \text{Exp}(\lambda) \\ \mathcal{N}(\mu, \sigma^2) \end{cases}$$

Markov Chains

$\pi_t = [\pi_t(1) \ \ \pi_t(2) \ \cdots \ \pi_t(n)] = [P(X_t=1) \ \ P(X_t=2) \cdots P(X_t=n)]$

represents distribution at time t

$\pi_0 =$ initial distribution

$$\pi_{t+1} = \pi_t P \implies \pi_n = \pi_0 P^n$$

$P_{i,j}$ = transition prob. from i to j

invariant distribution: $\pi = \pi P$

Irreducibility
("reachability")

Aperiodicity
(calculate and check $d(i) = 1$)
$\uparrow$
gcd { periods of state i }

Conditions of convergence ("Big Theorem")

Derivation of $\pi_{t+1} = \pi_t P$

$\pi_t = [\pi_t(0) \ \ \pi_t(1) \ \ \pi_t(2)]$

$\pi_{t+1}(i) = \pi_t(0) \cdot P_{0,i} + \pi_t(1) P_{1,i} + \pi_t(2) P_{2,i}$

$$\underbrace{[\pi_t(0) \ \ \pi_t(1) \ \ \pi_t(2)]}_{\pi_t} \underbrace{\begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} \\ P_{1,0} & P_{1,1} & P_{1,2} \\ P_{2,0} & P_{2,1} & P_{2,2} \end{bmatrix}}_{P}$$

$= \underbrace{[\pi_{t+1}(0) \ \ \pi_{t+1}(1) \ \ \pi_{t+1}(2)]}_{\pi_{t+1}}$

$\pi_t(0) \cdot P_{0,0} + \pi_t(1) P_{1,0} + \pi_t(2) P_{2,0}$

$\pi_t(0) \cdot P_{0,1} + \pi_t(1) \cdot P_{1,1} + \pi_t(2) \cdot P_{2,1}$

$\pi_t(0) P_{0,2} + \pi_t(1) \cdot P_{1,2} + \pi_t(2) \cdot P_{2,2}$