Basic Probability

Joint distribution-

$\mathbb{P}(x,y) = \mathbb{P}(X = x \,\&\, Y = y)$

$\mathbb{P}(x,y) = \mathbb{P}(x)\mathbb{P}(y) \Leftrightarrow$

X, Y *are independent.*

Conditional Probability (p(x) given y)-

$$\mathbb{P}(x|y) = \frac{\mathbb{P}(x,y)}{\mathbb{P}(y)}$$

$\mathbb{P}(x|y) = \mathbb{P}(x) \Leftrightarrow$

X, Y *are independent.*

Multivariable conditioning-

$\mathbb{P}(x,y|z) = \mathbb{P}(x|z)\mathbb{P}(y|z);$

$\mathbb{P}(x|z) = \mathbb{P}(x|y,z);\ \mathbb{P}(y|z) = \mathbb{P}(y|x,z);$

Chain rule-

$\mathbb{P}(x,y) = \mathbb{P}(x|y)\mathbb{P}(y) = \mathbb{P}(y|x)\mathbb{P}(x)$

Marginalization הסתברות השלמה-

- Discrete:

$\mathbb{P}(x) = \sum_y \mathbb{P}(x,y) = \sum_y \mathbb{P}(x|y)\mathbb{P}(y)$

- Continuous:

$\mathbb{P}(x) = \int_y \mathbb{P}(x,y) = \int_y \mathbb{P}(x|y)\mathbb{P}(y)$

Bayes Rule-

$$\boxed{\mathbb{P}(x|y) = \frac{\mathbb{P}(y|x)\mathbb{P}(x)}{\mathbb{P}(y)}}$$

- Discrete:

$$\mathbb{P}(x|y) = \frac{\mathbb{P}(y|x)\mathbb{P}(x)}{\sum_{x'} \mathbb{P}(y|x')\mathbb{P}(x')}$$

- Continuous:

$$\mathbb{P}(x|y) = \frac{\mathbb{P}(y|x)\mathbb{P}(x)}{\int_{x'} \mathbb{P}(y|x')\mathbb{P}(x')dx'}$$

additional variable/data:

$$\mathbb{P}(x|y,z) = \frac{\mathbb{P}(y|x,z)\mathbb{P}(x|z)}{\mathbb{P}(y|z)}$$

Expectation-

- Discrete: $\mathbb{E}[X] = \sum_i x_i \mathbb{P}(x_i)$
- Continuous: $\mathbb{E}[X] = \int x\mathbb{P}(x)dx$

$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

Covariance-

$$Cov[X] = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$$

Scalar case:

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

Gaussian distributions

Covariance form $x \sim \mathcal{N}(\mu, \Sigma)$

- 1-D pdf (mean=μ, var=σ^2):

$$\mathbb{P}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

- multi-dimensional pdf (mean=μ, var=σ^2):

$$\mathbb{P}(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}\|x - mu\|_\Sigma^2\right)$$

$$\boxed{\|a\|_\Sigma^2 = a^T\Sigma^{-1}a}$$

Information form-$x \sim \mathcal{N}(\eta, \Lambda)$

$\Lambda \doteq \Sigma^{-1};\ \eta = \Lambda\mu$

pdf:

$$\mathbb{P}(x) = \mathcal{N}^{-1}(\eta, \Lambda) = \frac{\exp\left(\frac{1}{2}\eta^T\Lambda^{-1}\eta\right)}{\sqrt{\det(2\pi\Lambda^{-1})}} \exp\left(-\frac{1}{2}x^T\Lambda x + \eta^T x\right)$$

Marginalization	Conditioning
$p(x) = \int p(x,y)\,dy \doteq \mathcal{N}(\mu, \Sigma) \doteq \mathcal{N}^{-1}(\eta, I)$	$p(x\|y) = \frac{p(x,y)}{p(y)} \doteq \mathcal{N}(\mu', \Sigma') \doteq \mathcal{N}^{-1}(\eta', I')$
$\mu = \mu_x$ $\Sigma = \Sigma_{xx}$	Cov form $\mu' = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$ $\Sigma' = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$
$\eta = \eta_x - I_{xy}I_{yy}^{-1}\eta_y$ $I = I_{xx} - I_{xy}I_{yy}^{-1}I_{yx}$	$\eta' = \eta_x - I_{xy}y$ $I' = I_{xx}$ Info form

Bayesian Inference

- $\mathbb{P}(x)$- **prior prob.** Knowledge regarding x **before** incorporating senor readings (z).
- $\mathbb{P}(x|z)$- **posterior prob.** Knowledge regarding x **after** incorporating senor readings (z).

Bayes rule-

$$\mathbb{P}(x|z) = \frac{\mathbb{P}(Z|X)\mathbb{P}(x)}{\mathbb{P}(z)} = \eta\mathbb{P}(z|x)\mathbb{P}(z)$$

Bayesian update-

$$\mathbb{P}(x|z_1, z_2, \ldots) = \frac{\mathbb{P}(z_n|x)\mathbb{P}(x|z_1, \ldots, z_{n-1})}{\mathbb{P}(z_n|z_{1,2}, \ldots z_{n-1})}$$

$= \eta_{1:n} \prod_{i=1}^n \mathbb{P}(z_i|x)\mathbb{P}(x)$

Motion model (Transition model)-

$x_{k+1} \sim \mathbb{P}(x_{k+1}|x_k, a_k) = f(x_k, a_k, w_k)$

Observation model-

$z_k \sim \mathbb{P}(z_k|x_k) = h(x_k, v_k)$

For given models, the only reason for stochasticity is the noise (w_k, v_k).

Markov assumptions-

$\mathbb{P}(x_k|x_{0:k-1}, z_{1:k-1}, a_{0:k-1}) = \mathbb{P}(x_k|x_{k-1}, a_{k-1})$

$\mathbb{P}(z_k|x_{0:k}, a_{0:k-1}, z_{1:k-1}) = \mathbb{P}(z_k|x_k)$

Gaussian noise-

$$\mathbb{P}(x_{k+1}|x_k, a_k) = \frac{\exp\left(-\frac{1}{2}\|x_{k+1} - f(x_k, a_k)\|_{\Sigma_W}^2\right)}{\sqrt{\det(2\pi\Sigma_W)}}$$

$$\mathbb{P}(z_k|x_k) = \frac{\exp\left(-\frac{1}{2}\|z_k - h(x_k)\|_{\Sigma_v}^2\right)}{\sqrt{\det(2\pi\Sigma_v)}}$$

Posterior belief-

$b[x_k] \doteq \mathbb{P}(x_k|a_{0:k-1}, z_{1:k}) \equiv \mathbb{P}(x_k|H_k)$

$\overline{b[x_k]} \doteq \mathbb{P}(x_k|a_{0:k-1}, z_{1:k-1}) \equiv \mathbb{P}(x_k|H_k^-)$

Environment Representation

Standard occupancy grid:

$\mathbb{P}(m|z_{1:k}, x_{1:k}) = \prod_i \mathbb{P}(m_i|z_{1:k}, x_{1:k})$

Occupancy grid – Binary bayes filter(static):

$b[x_k] = p(x|z_{1:k}, a_{1:k}) = p(x|z_{1:k})$

MDP, POMDP

MDP Objective-

$J(X_0, a_{0:T-1}) \doteq \mathbb{E}_X\{\sum_{t=0}^{T-1} r(X_t, a_t) + r_T(X_T)\}$

MDP Value function-

- Finite horizon:

$V_0^\pi(X) \doteq \mathbb{E}_X[\sum_{t=0}^{T-1} r(X_t, a_t) + r_T(X_T)|$
$X_0 = X, a_t = \pi_t(X_t)]$

- Discounted infinite horizon:

$V_0^\pi(X) \doteq \mathbb{E}_X[\sum_{t=0}^{\infty} \gamma^t r(X_t, a_t)\,|X_0 = X, a_t = \pi_t(X_t)]$

*$\gamma \to 0$: greedy

*$\gamma \to 1$: long horizon

POMDP Value function

- Finite Horizon:

$V_0^\pi(b_0) \doteq \mathbb{E}_z\{\sum_{l=0}^{L-1} r(b[X_l], a_l)|a_l = \pi(b[X_l])\}$

- Discounted Infinite Horizon:

$V_0^\pi(b_0) \doteq \mathbb{E}_z\{\sum_t \gamma^t r(b_t, a_t)\,|a_t = \pi(b_t)\}$

Belief MDP update

$$b_k[X_k] = \frac{\mathbb{P}_Z(z_k|X_k)\int_{x_{k-1}} \mathbb{P}_T(X_k|X_{k-1}, a_{k-1})b[X_{k-1}]dX_{k-1}}{\mathbb{P}(z_k|H_{k-1}, a_{k-1})}$$

Transformed transition model-

$\mathbb{P}_\psi(b_{k+1}|b_k, a_k) =$
$\mathbb{E}_{z_{k+1} \sim \mathbb{P}(\cdot|b_k, a_k)}[\mathbb{P}(b_{k+1}|b_k, a_k, z_{k+1})]$
$= \int_{z_{k+1}} \mathbb{P}(z_{k+1}|b_k, a_k)\mathbf{1}[b_{k+1} = \psi(b_k, a_k, z_{k+1})]dz_{k+1}$

*Open loop: actions are determined at once at time 0 ($f(X_0)$ *or* b_0).

*Close loop: actions are determined "just-in-time" ($f(X_t)$*or* H_k).

Belief Space Planning

Objective function-

$J(b[X_k], a_{k:k+L-1}) =$
$\mathbb{E}_{z_{k+1:k+L}}\{\sum_{l=0}^{L-1} r(b[X_{k+l}], a_{k+l}) + r(b[X_{k+L}])\} =$
$r(b_k, a_k) + \int \mathbb{P}(z_{k+1}|H_k, a_k)\big[r(b_{(k+1)}, a_{k+1}) +$
$\int \mathbb{P}(z_{k+2}|H_{k+1}, a_{k+1})[r_2 + \cdots]dz_{k+2}\big]dz_{k+1}$

* Weighted rewards (multi-Objective)-

$r(b,a) = \boldsymbol{w} \cdot \boldsymbol{r}(b,a)$

Continuous action space-

$$\nabla_a J(b_k, a) = \frac{\partial}{\partial a}\mathbb{E}\{\sum_{l=0}^{L-1} r(b_{k+l}, a_{k+l}) + r(b_{k+L})\} \Rightarrow \frac{\partial J}{\partial a_{i,j}} \approx \frac{\left(J(b_k, a + \epsilon\mathbf{1}_{(i,j)}) - J(b_k, a)\right)}{\epsilon}$$

Belief propagation-

$b[X_{k+l}] = \mathbb{P}(X_{k+l}|H_k, a_{k:k+l-1}, z_{k+1:k+l})$

- smoothing:

$b[X_{0:k+l}] =$
$b[X_{0:k}]\prod_{i=k+1}^{k+l} \mathbb{P}(X_i|X_{i-1}, a_{i-1})\mathbb{P}(z_i|X_i)$

- recursive:

$b[X_{0:k+l}] = \int_{X_{k:k+l-1}} b[X_k]$
$\prod_{i=k+1}^{k+l} \mathbb{P}(X_i|X_{i-1}, a_{i-1})\mathbb{P}(z_i|X_i)\,dX_{xk:k+l-1}$

Probabilistic Inference (Gausian Distributions)-

problem statement: $b[X_k] \sim \mathcal{N}(xX_k^*, \Sigma_k)$,
$z = h(x) + v,\ v \sim N(0, \Sigma_v)$. Need to solve
$X_{k:k+L}^* = \arg\max_{X_{k:k+L}} b[X_{k:k+L}] =$
$\arg\min_{X_{k:k+L}} -\log b[X_{k:k+L}] =$
$\arg\min_{X_{k:k+L}} \left(\|X_k - X_k^*\|_{\Sigma_k}^2 + \sum_{l=k+1}^{k+L}\|X_l - f(X_{l-1}, a_{l-1})\|_w^2 + \|z_l - h(X_l)\|_{\Sigma_v}^2\right)$

Solution:

1. Linearization (1st Order Taylor) $X = \bar{X} + \Delta X$

*Mahalanobis norm: $\|a\|^2{}_{\Sigma_v} = \left\|\Sigma_v^{-\frac{1}{2}}a\right\|^2$

2. Solve $A\Delta X = b$:

$$A_i = \begin{bmatrix} 0 & 0 \ldots & \underbrace{-\Sigma_w^{-\frac{1}{2}}\nabla_X f}_{\Delta X_{l-1}} & \Sigma_w^{-\frac{1}{2}} & 0 & \ldots & 0 \end{bmatrix}$$

$b_i = \Sigma_w^{-\frac{1}{2}}\left(f(\overline{X}_{l-1}, a_{l-1}) - \overline{X}_l\right)$

$\Delta X_{k:k+L}^* = \arg\min_{\Delta X_{k+L}} \|\mathcal{A}\Delta X_{k:k+L} - b\|^2$

$\to \Delta X_{k:k+L}^* = (\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T b$

$X_{k:k+1}$ case:

$$A = \begin{bmatrix} \Sigma_k^{-\frac{1}{2}} & 0 \\ -\Sigma_w^{-\frac{1}{2}}F & \Sigma_w^{-\frac{1}{2}} \\ 0 & \Sigma_v^{-\frac{1}{2}} \end{bmatrix}, b = \begin{bmatrix} -\Sigma_k^{-\frac{1}{2}}\mu_k \\ -\Sigma_w^{-\frac{1}{2}}a_k \\ -\Sigma_v^{-\frac{1}{2}}z_{k+1} \end{bmatrix}$$

3. Update until convergence:

$\overline{X}_{k:k+L} \leftarrow \overline{X}_{k:k+L} + \Delta X_{k:k+L}^*$

$b[X_{k:k+L}] \equiv \mathbb{P}(X_{k:k+L}|H_k, a_{k:k+L-1}, z_{k+1:k+L})$
$= \mathcal{N}(X_{k:k+L}^*, \mathcal{A}^T\mathcal{A})$

Maximum likelihood-

$z_{k+l} \sim \mathbb{P}(z_{k+l}|\overline{X}_{k+1}),\ z_{k+1} = h(\overline{X}_{k+1}) + v$

Kalman update: $\bar{x}^+ = \bar{x}^- + K\left(z - h(\bar{x}^-)\right)$

Zero innovation = no noise

Information Theoretic Costs

Entropy-

- Discrete: $\mathcal{H}[X] = -\sum_{i=1}^n \mathbb{P}(x_i)\log\mathbb{P}(x_i) = \mathbb{E}_{X \sim \mathbb{P}(X)}\{-\log\mathbb{P}(x_i)\} \geq 0$
- Continuous: $\mathcal{H}[X] = -\int_{\mathbb{X}} \mathbb{P}(x)\log\mathbb{P}(x)\,dx = \mathbb{E}_{X \sim \mathbb{P}(X)}\{-\log\mathbb{P}(x_i)\}$

Joint Entropy- measure of uncertainty

- Discrete:

$\mathcal{H}[X,Y] = \mathbb{E}_{X,Y \sim \mathbb{P}(X,Y)}\{-log\mathbb{P}(X,Y)\}$

- Continuous:

$\mathcal{H}[X,Y] = -\int_{\mathbb{X},\mathbb{Y}} \mathbb{P}(x,y)log\mathbb{P}(x,y)dxdy$

Conditional Entropy-

$\mathcal{H}[Y|X] = \mathbb{E}_x\{\mathcal{H}[Y|X = x]\} =$
$-\mathbb{E}_{X,Y \sim \mathbb{P}(X,Y)}\left\{\log\frac{\mathbb{P}(X,Y)}{\mathbb{P}(X)}\right\} =$
$-\mathbb{E}_{X,Y \sim \mathbb{P}(X,Y)}\{log\mathbb{P}(Y|X)\} = \Sigma_{x,y}p(x,y)\log\left(\frac{p(x)}{p(x,y)}\right)$

*$\mathcal{H}[X,Y] = \mathcal{H}[X] + \mathcal{H}[Y|X]$

* if X,Y independent $\mathcal{H}[X,Y] = \mathcal{H}[X] + \mathcal{H}[Y]$

*$\mathcal{H}[X,Y] \leq \mathcal{H}[X] + \mathcal{H}[Y]$

*$\mathcal{H}[Y|X] \leq \mathcal{H}[Y]$

Mutual Information- reduction in the uncertainty of X from knowing Y.

MI expresses the amount of information shared between a set of variables

$MI[X;Y] = \mathcal{H}[X] - \mathcal{H}[X|Y]$

*$MI(X;Y) = MI(Y;X)$

*if X,Y independent: $MI(X;Y) = 0$

*$MI[X;Y|Z] = \mathcal{H}[X|Z] - \mathcal{H}[X|Y,Z]$

* Chain rule:

$MI[X_1, X_2, \ldots; Y] = \sum_{i=1}^n MI[X_i; Y|X_{<i}]$

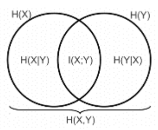
Information Gain- reduction in entropy over X given Y, assumes specific value of y.

$IG[X; Y = y] = \mathcal{H}[X] - \mathcal{H}[X|Y = y]$

*Relation to MI:

$$MI[X;Y] = \mathbb{E}_y\{IG[X;Y=y]\}$$

MI is the expected information gain.



Costs/Reward in BSP-

- Entropy:
$$\mathcal{H}[b[X_{k+l}]] = -\int_{X_{k+l}} b[X_{k+l}] \log b[X_{k+l}]\, dX_{k+l}$$
- IG: $\mathbb{E}_{z_{k+1}}\{r(b[X_{k+1}])\} = \mathbb{E}_{z_{k+1}}\{IG[X_{k+1}; Z_{k+1} = z_{k+1}]\} = MI[\mathbb{P}(X_{k+1}|H_k, a_k); Z_{k+1}]$

Gaussian models- Entropy-

$$\mathcal{H}[p(X)] = \frac{1}{2}\ln[(2\pi e)^N\, det(\Sigma)]$$

Entropy corresponds to the volume of the uncertainty ellipsoid.

Distances (Metrics)-

1. $d(x,y) \geq 0, \quad d(x,y) = 0 \leftrightarrow x = y,$
   $d(x,y) = d(y,x) ||| \; d(x,z) \leq d(x,y) + d(y,z)$

- L1 (Kolmogorov Variational)-
$$d_{L_1}(\mathbb{P},\mathbb{Q}) = \int_X |\mathbb{P}(X) - \mathbb{Q}(X)| dX$$
- L2 (Integral Square Error)-
$$d_{L_2}(\mathbb{P},\mathbb{Q}) = \int_X [\mathbb{P}(X) - \mathbb{Q}(X)]^2 dX$$

KL-Divergence

$$KL[\mathbb{P}||\mathbb{Q}] = \int_X \mathbb{P}(X)\log\frac{\mathbb{P}(X)}{\mathbb{Q}(X)} dX$$

$$= \mathbb{E}_{X\sim\mathbb{P}}\left\{\log\frac{\mathbb{P}(X)}{\mathbb{Q}(X)}\right\}$$

$$= \mathcal{H}[\mathbb{P}(X)] - \mathbb{E}_{X\sim\mathbb{P}}\{\log \mathbb{Q}(X)\} \geq 0$$

*if $\mathbb{P} \equiv \mathbb{Q} \rightarrow d_{KL} = 0$ $KL[\mathbb{P}(X,Y)||\mathbb{P}(X)\mathbb{P}(Y)]$

$$= MI[X;Y] \equiv \mathcal{H}[X] - \mathcal{H}[X|Y]$$

*not true metric

*Multi-variate Gaussian $X \in \mathbb{R}^N$

$$\mathbb{P}(X) = \mathcal{N}(\mu_1, \Sigma_1), \mathbb{Q}(X) = \mathcal{N}(\mu_2, \Sigma_2)$$

$$KL[\mathbb{P}||\mathbb{Q}] = \frac{1}{2}[tr(\Sigma_2^{-1}\Sigma_1) + (\mu_1 + \mu_2)^T \Sigma_1^{-1}(\mu_1 + \mu_2) - N + \ln\det\left(\frac{\Sigma_2}{\Sigma_1}\right)]$$

Jensen inequality-

For any convex $g(x)$: $g(\mathbb{E}\{X\}) \leq \mathbb{E}\{g(X)\}$

Jensen-Shannon Divergence-

$$JSD[\mathbb{P}||\mathbb{Q}] = KL[\mathbb{P}||\mathbb{M}] + KL[\mathbb{Q}||\mathbb{M}]$$

$$\mathbb{M} = \frac{\mathbb{P}+\mathbb{Q}}{2}$$

*not true metric

Search-based Planning

Algorithm 3 Label Correcting Algorithm

```
OPEN ← {s}, g_s = 0, g_i = ∞ for all i ∈ V \ {s}
while OPEN is not empty do
    Remove i from OPEN
    for j ∈ Children(i) do
        if (g_i + c_ij) < g_j and (g_i + c_ij) < g_τ then
            g_j ← (g_i + c_ij)
            Parent(j) ← i
            if j ≠ τ then
                OPEN ← OPEN ∪{j}
```

▷ Only when $c_{ij} \geq 0$ for all $i, j \in \mathcal{V}$

* BFS- first in, first out (פיתוח רוחבי)
* DFS- last in, first out (פיתוח לעומק)
* Dijkstra- priority queue (פיתוח לפי תג מינימלי)
A- Informed (heuristic) ((מרחק מההתחלה ומרחק אל היעד

Heuristic function-

1. Admissible- $h_i \leq dist(i, goal) \forall\, i \in \mathcal{V}$
2. Consistent- triangle inequality- $h_{goal} = 0,$
$h_i \leq c_{ij} + h_j\ \forall i \neq goal,\ j \in Children(i)$
3. ϵ-Consitent- $h_{goal} = 0, h_i \leq \epsilon c_{ij} + h_j$
$\forall i \neq goal,\ j \in Children(i)$

* if h^1, h^2 consistent:
1. $h = \max\{h^1, h^2\}$ is consistent
2. $h = h^1 + h^2$ is ϵ-consistent ($\epsilon = 2$)

Sampling-based Planning

*RRT- sample rand node x_{rand}, steer towards x_{rand} fixed d to get x_j. if the route from the closest x_i is collision free, add x_j into the tree.

Dynamic constraints- sample controls to generate feasible paths.

RRT- sample rand node x_{rand}, find all nodes in its region. Connect to node with the known shortest path from the start.

LQG-

Linear model, Quadric costs, Gaussian noise.

$$X_{k+1} = AX_k + Ba_k + G_{w_k}, \qquad w_k \sim \mathcal{N}(0, \Sigma_w)$$
$$z_k = HX_k + v_k, \qquad v_k \sim \mathcal{N}(0, \Sigma_v)$$
$$b_{k+l} = \mathcal{N}(\hat{X}_{k+l}, \Sigma_{k+l})$$
$$J_k(b_k, a_k) = \mathbb{E}\{\textstyle\sum_{l=0}^{L-1} \|X_{k+l} - \xi_{k+l}\|^2_{W_X} + \|a_{k+l}\|^2_{W_a} \,|\, a_{k+l} = \pi(b_{k+l})\}$$

Solution for l=1 (Kalman Filter):

$$b_{k+1} = \begin{bmatrix}\hat{X}_{k+1} \\ \Sigma_{k+1}\end{bmatrix} = \begin{bmatrix} A\hat{X}_k + Ba_k + K_{k+1}\left(z_{k+1} - H\left(A\hat{X}_k + Ba_k\right)\right) \\ (I - K_{k+1}H)(A\Sigma_k A^T + G\Sigma_w G^T) \\ with\ \ K_{k+1} = \Sigma_k A^T (A\Sigma_k A^T + \Sigma_v)^{-1}\end{bmatrix}$$

Optimal policy:

$$a_{k+l} = \pi(b_{k+l}) = -K_{k+l}\hat{X}_{k+l}$$

MDP – Approachs

Value Function-

- Finite Horizon-
$$V_0^\pi(X_0) = \mathbb{E}\left[\sum_{t=0}^{T-1} \begin{matrix} r(X_t, a_t) + r_{T(X_T)} \\ |a_t = \pi_t(X_t)\end{matrix}\right] = r(X, \pi(X)) + \mathbb{E}_{X' \sim \mathbb{P}_T(X'|X, \pi_0(X))}\{V_1^\pi(X')\}$$
- Discounted infinite Horizon-
$$V^\pi(X) = \mathbb{E}\left[\sum_{t=0}^{\infty} \begin{matrix}\gamma^t r(X_t, a_t) \\ |X_0 = X, a_t = \pi(X_t)\end{matrix}\right] = r(X, \pi(X)) + \mathbb{E}_{X' \sim \mathbb{P}_T(X'|X, \pi(X))}\{V^\pi(X')\}$$
$$V^*(X) = \max_\pi V^\pi(X)$$
$$\pi^*(X) = \arg\max_\pi V^\pi(X)$$

Bellman equation:

$$V^*(X) = \max_{a\in\mathcal{A}}\left[r(X,a) + \gamma \sum_{X'\in\mathcal{X}} \mathbb{P}_{T(X'|X,a)} V^*(X)\right]$$

Q Function:

$$Q^\pi(X,a) = \mathbb{E}\left\{\sum_{t=0}^{\infty} \begin{matrix}\gamma^t r(X_t, a_t)| \\ X_0 = X, a_0 = a, a_t = \pi(X_t)\end{matrix}\right\}$$
$$= r(X,a) + \gamma \mathbb{E}_{X' \sim \mathbb{P}_T(\cdot|X,a)}\{V^\pi(X')\}$$
$$= r(X,a) + \gamma \mathbb{E}_{X' \sim \mathbb{P}_T(\cdot|X,a)}\{Q^\pi(X', \pi(X'))\}$$

*Determinstic Policy:

$$V^\pi(X) = Q^\pi(X, \pi(X))$$

*Stochastic Policy:

$$V^\pi(X) = \sum_{a\in\mathcal{A}} \pi(a|X) Q^\pi(X,a)$$

Value Iteration – Infinite Horizon-

$$V_{n+1}(X) = \max_{a\in\mathcal{A}}\left[r(X,a) + \gamma \sum_{X'\in\mathcal{X}} \mathbb{P}_T(X'|X,a) V_n(X')\right]$$
$$V_{n\to\infty}(X) = V^*(X)$$

Policy Iteration – Infinite Horizon-

1. Policy Evaluation:
$$V^\pi(X) = \mathbb{E}\left[\sum_{t=0}^{\infty} \begin{matrix}\gamma^t r(X_t, a_t)| \\ X_0 = X, a_t = \pi(X_t)\end{matrix}\right]$$
$$V_{n+1}(X) = \max_{a\in\mathcal{A}}\left[r(X, \pi(X)) + \gamma \sum_{X'\in\mathcal{X}} \mathbb{P}_T(X'|X, \pi(X)) V_n(X')\right]$$
2. Policy Improvement:
$$\pi'(X) = \arg\max_{a\in\mathcal{A}} Q^\pi(X,a) \equiv \arg\max_{a\in\mathcal{A}}\left\{r(X,a) + \gamma \sum_{X' \in script X} \mathbb{P}_T(X'|X,a) V^\pi(X')\right\}$$

Bellman Operator-

Infinity norm-

$$\|U - V\|_\infty = \max_{X\in\mathcal{X}} |U(X) - V(X)|$$
$$T^\pi[V](X) = r(X, \pi(X)) + \gamma \mathbb{E}_{X' \sim \mathbb{P}_T(X'|X, \pi(X))}\{V(X')\}$$
$$= \boldsymbol{r}^\pi + \gamma \mathcal{P}^\pi V$$
$$*T^*[V] = V$$

gamma contraction-

$$\|T^*[U] - T^*[V]\|_\infty \leq \gamma \|U - V\|_\infty$$