Monte Carlo approximation :

→ distribution of $f(rv)$ - difficult

→ So, generate S samples $= \{x_1, \dots, x_S\}$

find distribution $f(X)$

$$\{f(x_s)\}_{s=1}^{S}$$

→ $\mathbb{E}[f(X)] = \int f(x)\, p(x)\, dx \simeq \frac{1}{S} \sum_{s=1}^{S} f(x_s)$

$$x_s \sim p(x)$$

- $\bar{x} = \frac{1}{S} \sum_{s=1}^{S} x_s \longrightarrow \mathbb{E}[x]$
- $\frac{1}{S} \sum_{s=1}^{S} (x_s - \bar{x})^2 \longrightarrow \mathrm{var}[x]$
- $\mathrm{median}\{x_1, \dots, x_S\} \longrightarrow \mathrm{median}(x)$

→ The accuracy of M.C increases with sample size

→ Approximat of M.C.

if exact mean $\cdot\ \mu = \mathbb{E}[f(x)]$

MC $\hat{\mu}$ is .

$$(\hat{\mu} - \mu) \longrightarrow N\left(0, \frac{\sigma^2}{S}\right)$$

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

$$\hat{\sigma}^2 = \frac{1}{S} \sum_{s=1}^{S} (f(x_s) - \hat{\mu})^2$$

Then,

$$P\left\{\mu - 1.96 \frac{\hat{\sigma}}{\sqrt{S}} < \hat{\mu} \leq \mu + 1.96 \frac{\hat{\sigma}}{\sqrt{S}}\right\} \approx 0.95$$

$\therefore\ S \geq \frac{4\hat{\sigma}^2}{\epsilon^2}$ [$\pm\epsilon$. ans which is accurate to within $\pm\epsilon$]

Entropy:

$H(X)$ or $H(p)$

$$H(X) \triangleq - \sum_{k=1}^{K} p(X=k) \log_2 p(X=k)$$

$\log_{2}$ = Bits (short of binary digits)

$\log_{e}$ = nats {natural digits}

KL- divergence

- Dissimilarity of 2- prob. distribution, p, q

$$KL(p \| q) \triangleq \sum_{k=1}^{K} p_k \log \frac{p_k}{q_k}$$

$$= -H(p) + H(p, q)$$

$$H(p, q) \triangleq - \sum_{k} p_k \log q_k \quad \text{[Cross Entropy]}$$

Mutual information:

- how much knowing one variable tells about other variable

~ option correlation

$$\mathbb{I}(X; Y) \triangleq KL\left(p(X, Y) \| p(X) p(Y) \right)$$

$$= \sum_{x} \sum_{y} p(x, y) \log \frac{p(x, y)}{p(x)\, p(y)}$$

Next: Generative models for discrete data