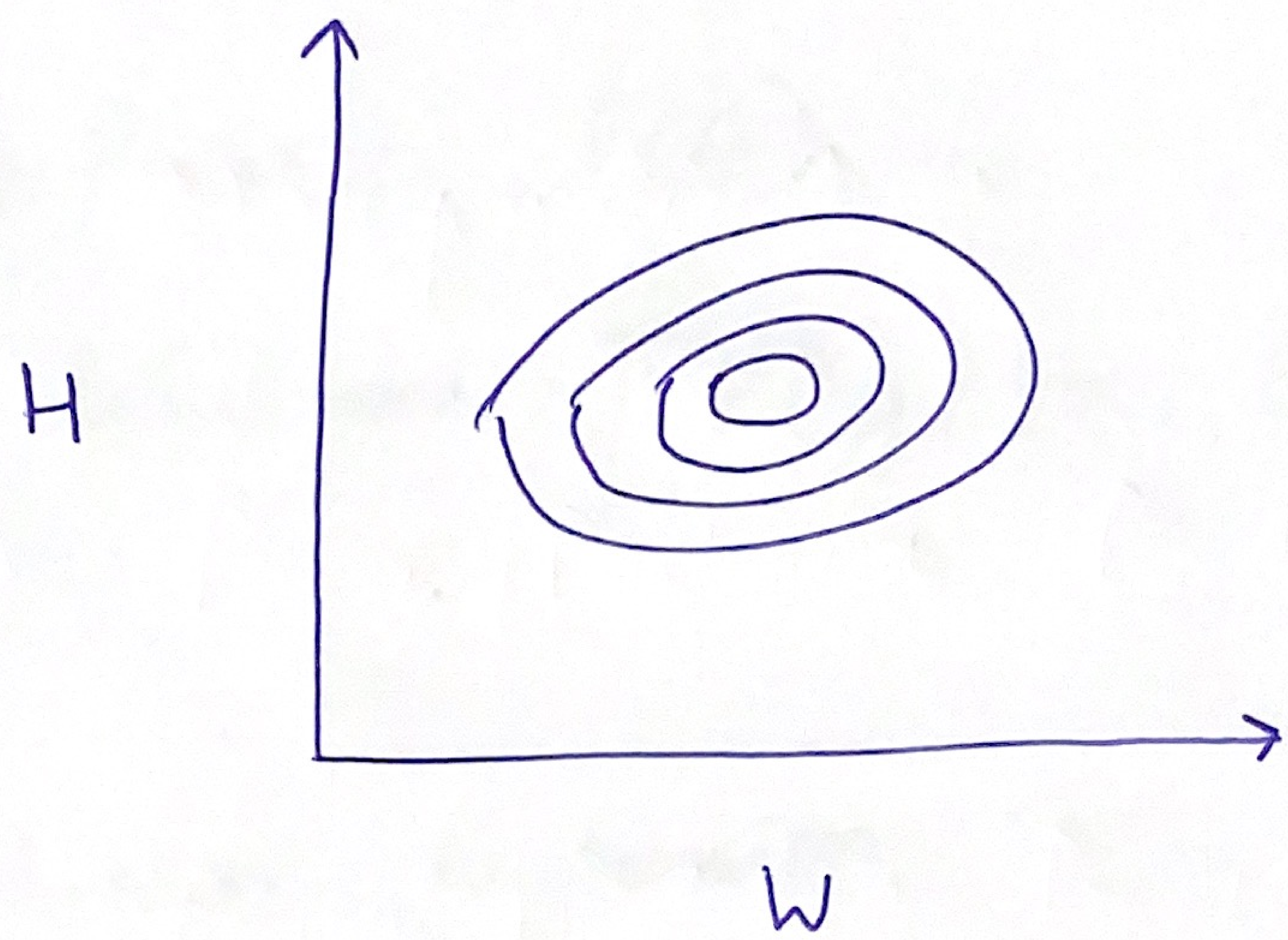
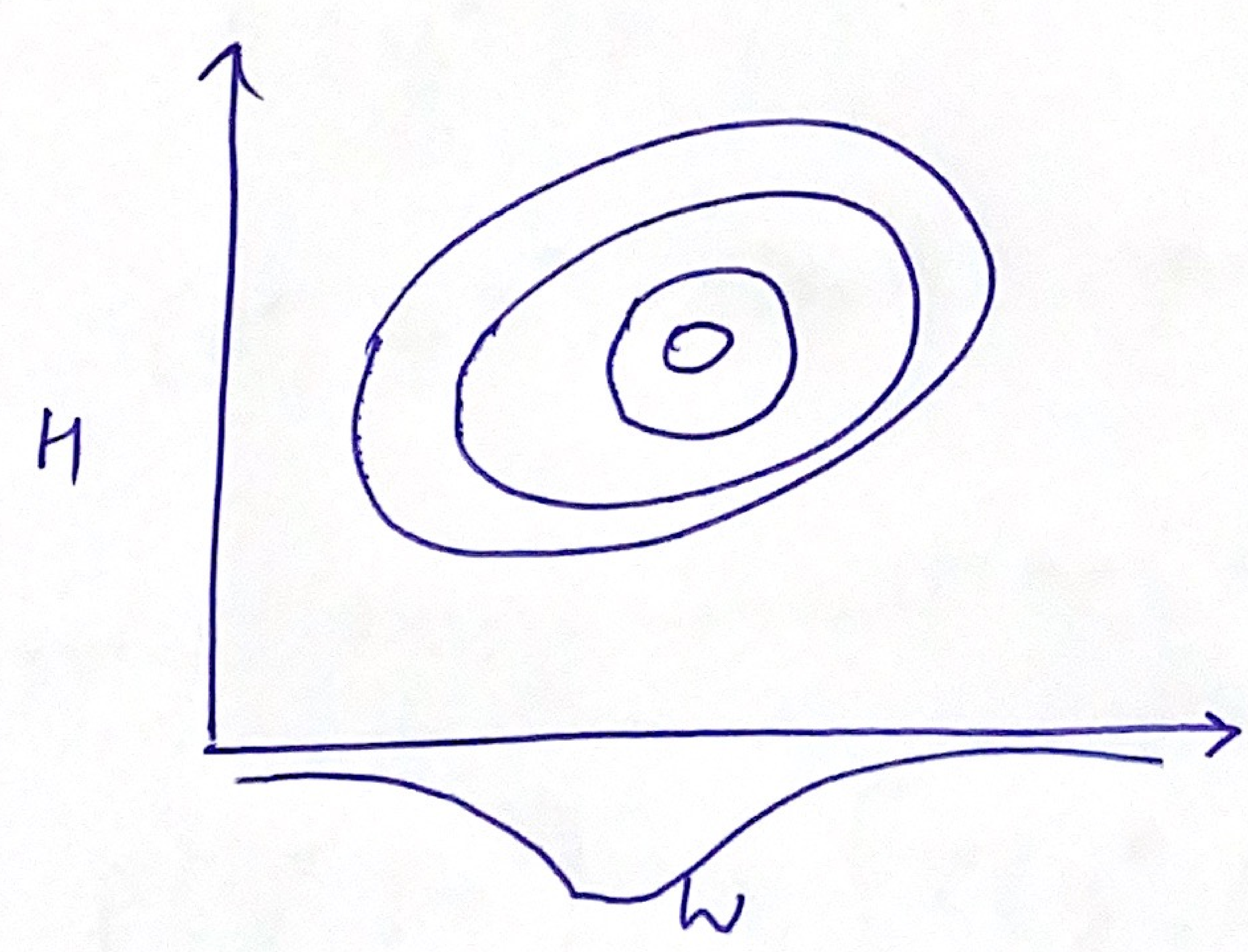
# Marginalisation




$$P(H,W) = N_2(\mu, \Sigma)$$

$$P(H) = ?$$

$$= \int_{W_L}^{W_H} P(H,W)\, dW$$

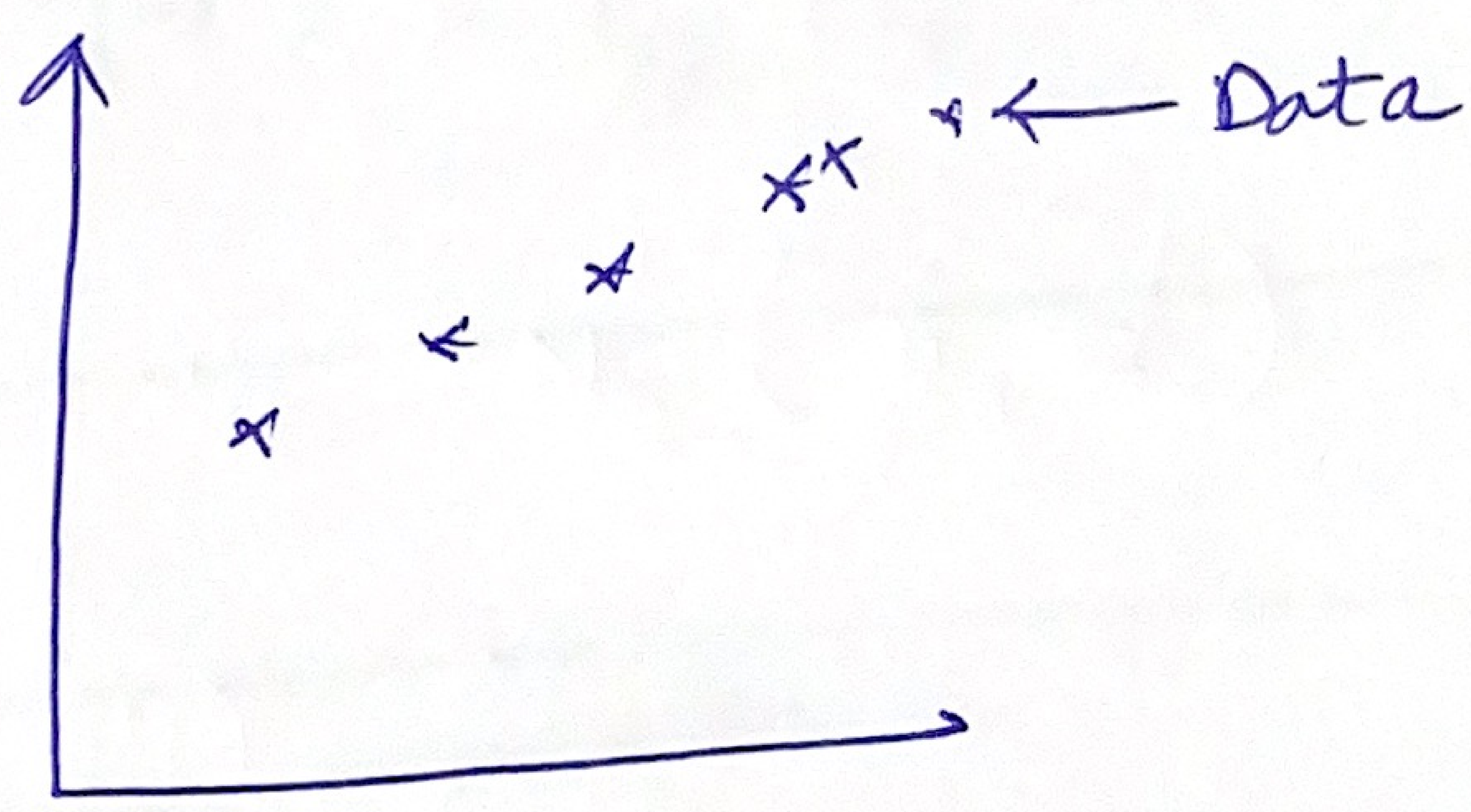


# Prior Predictive Distribution.

Prior : $P(\theta)$

Likelihood : $\prod_{i=1}^{N} P(y_i \mid x_i, \theta)$

## Linear Regression




Q) Functions from prior look like?

Or). $P(y^* \mid x^*) = ?$

$(x^*, y^*)$ : Test input and output.

Apply Marginalisation (Prior distribution only — No Data!)

$$P(y^* \mid x^*) = \int P(y^*, \theta \mid x^*) \, d\theta$$

$$P(y^*|x^*) = \int P(y^*, \theta | x^*)\, d\theta$$

Apply Bayes Rule $\quad P(A,B) = P(A|B) \cdot P(B)$

$$P(y^*|x^*) = \int P(y^*|\theta, x^*)\, P(\theta)\, d\theta$$

PRIOR PREDICTIVE DISTRIBUTION

$$P(y^*|x^*) = \mathbb{E}_{P(\theta)}\left[P(y^*|\theta, x^*)\right]$$

For linear regression

Assume

$P(\theta) = N(m_0, S_0)$ ; likelihood noise $= \sigma^2$

$$P(y^*|x^*) = N\left(x^{*T} m_0,\ x^{*T} S_0 x^* + \sigma^2\right)$$

# POSTERIOR PREDICTIVE DISTRIBUTION

Prior: $P(\theta)$

Data: $D = \{(x_1, y_1) \dots (x_N, y_N)\}$

Posterior $= P(\theta | D)$

Q) $p(y^* | x^*, D) = ?$

Marginalize out $\theta$

<u>BUT</u> $\theta$ comes from Posterior

$$p(y^* | x^*, D) = \int p(y^*, \theta | x^*, D)\, d\theta$$

Apply Bayes Rule

$$p(y^* | x^*, D) = \int p(y^* | x^*, D, \theta)\, p(\theta | x^*, D)\, d\theta$$

1) But $y^*$ does not depend on $D$ if $\theta$ is known

2) $\theta$ does not depend on $x^*$

$$P(y^*|x^*, D) = \int P(y^*|x^*, \theta) . P(\theta|D)\, d\theta$$

Posterior

Posterior Predictive Distribution

$$P(y^*|x^*, D) = E_{P(\theta|D)}\left[P(y^*|x^*, \theta)\right]$$

Posterior

Linear regression

$$P(\theta|D) = N(m_N, S_N)$$

$$P(y^*|x^*, D) = N(x^{*T} m_N,\; x^{*T} S_N x^* + \sigma^2)$$

# Logistic Regression Posterior Predictive

Prior: $P(\theta) = N(\mu, \Sigma)$

Likelihood: Bernoulli $(\sigma(x_i^T \theta)$

Posterior (w/ Laplace Approximation)

$: N(\theta \mid \theta_{MAP}, H^{-1})$

where $H^{-1} = -\nabla^2 \log(D, \theta_{MAP})$

Posterior Predictive

$$p(y^* \mid x^*, D) = \int \underbrace{p(y^* \mid x^*, \theta)}_{\substack{\text{Bernoulli} \\ (\sigma(x^{*T}\theta))}} . \underbrace{P(\theta \mid D)}_{\text{Normal}} \, d\theta$$

INTRACTABLE!

# Monte-Carlo Sampling

- Solve integrals of form $I = \int f(x) p(x)\, dx$
- By sampling

$$\boxed{\int f(x)\, p(x) = E_p[f] = \frac{\sum_{i=1}^{S} f(x_i)}{S} \quad \text{where } x_i \sim p(x) \text{ i.i.d.}}$$

Q) Find mean of std. Normal. ($\mu = 0, \sigma = 1$)

$f(x) = x$

$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-0}{\sigma}\right)^2\right) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$

$$\hat{\mu} = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{x}_{\text{ODD}} \underbrace{\exp\left(-\frac{x^2}{2}\right)}_{\text{EVEN}} dx$$

= ODD FUNCTION

$\hat{\mu} = 0$

Using Monte Carlo Method.

$$\hat{\mu} = \frac{\sum_{i=1}^{S} x_i}{S}$$

$x_i \sim N(0,1)$ i.i.d.

Find std Variance of $x \sim N(\mu, \sigma)$

$f(x) = /\Sigma x_i$ $\quad x_i \sim N(\mu, \sigma)$

$$E(x^2) = \frac{\sum_{i=1}^{S} x_i^2}{S}$$

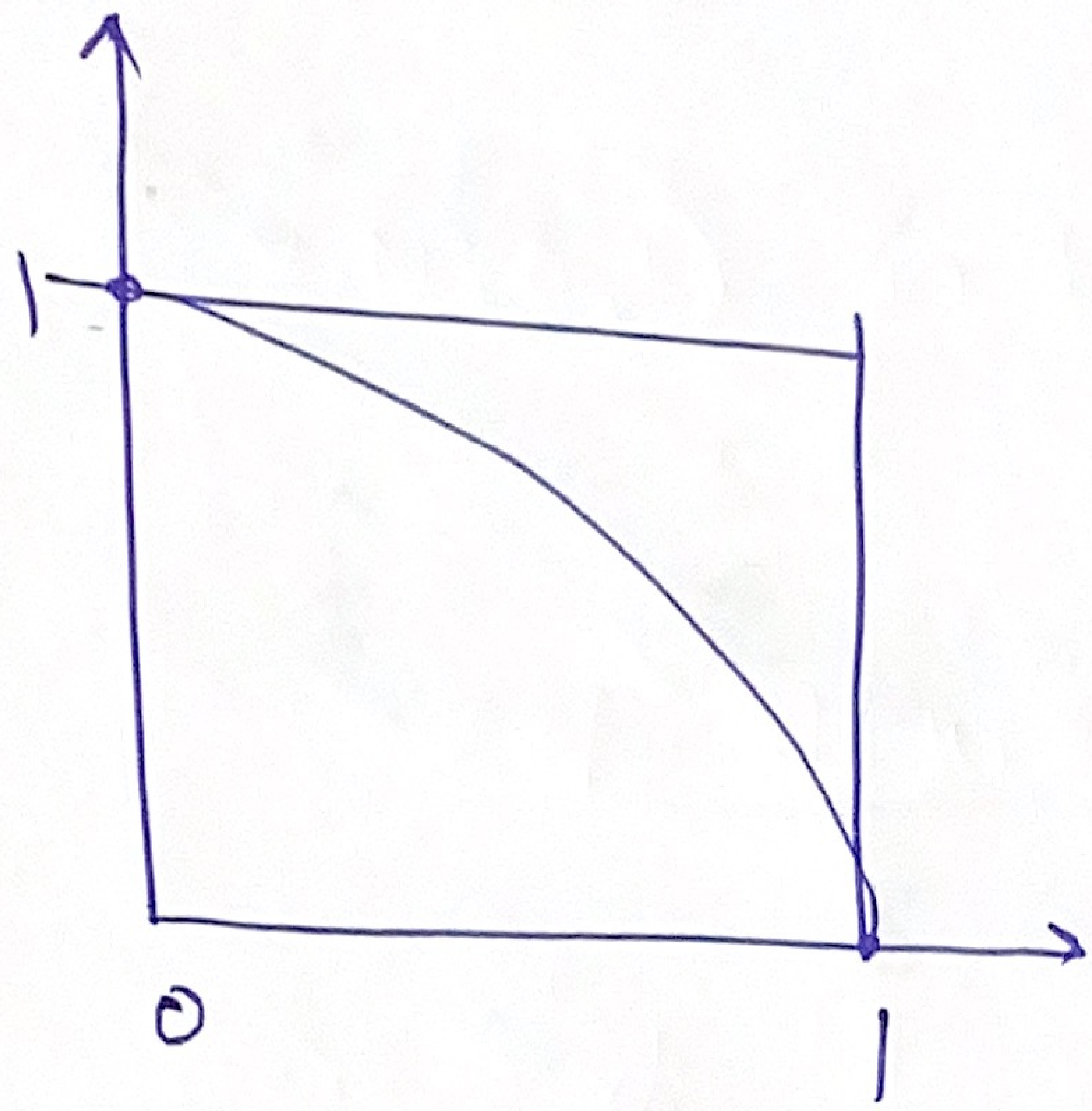
$$E(x) = \frac{\sum_{i=1}^{S} x_i}{S}$$

$$Var(x) = E(x^2) - (E(x))^2$$

Q) find value of $\pi$ using. M.C. Sampling

$f(x) = ?$

$p(x) = ?$



$$p(x) = U_2(0,1)$$

$\uparrow$ 2 dim.

$$f(x) = I(x^T x \le 1)$$

$$= \begin{cases} 1; & \text{if } x^T x \le 1 \\ 0; & \text{o/w} \end{cases}$$

$$\frac{\pi \times 1^2}{4} = E_p[f(x)] = \frac{\sum_{i=1}^{S} I(x_i^T x_i \le 1)}{S}$$

$$x_i \sim U_2(0,1)$$

# MC sampling - Unbiased Estimator

$$\phi = \int f(x)\, p(x)\, dx = E_p(f)$$

$x_i \sim p$; $i = 1, \ldots S$, i.i.d.

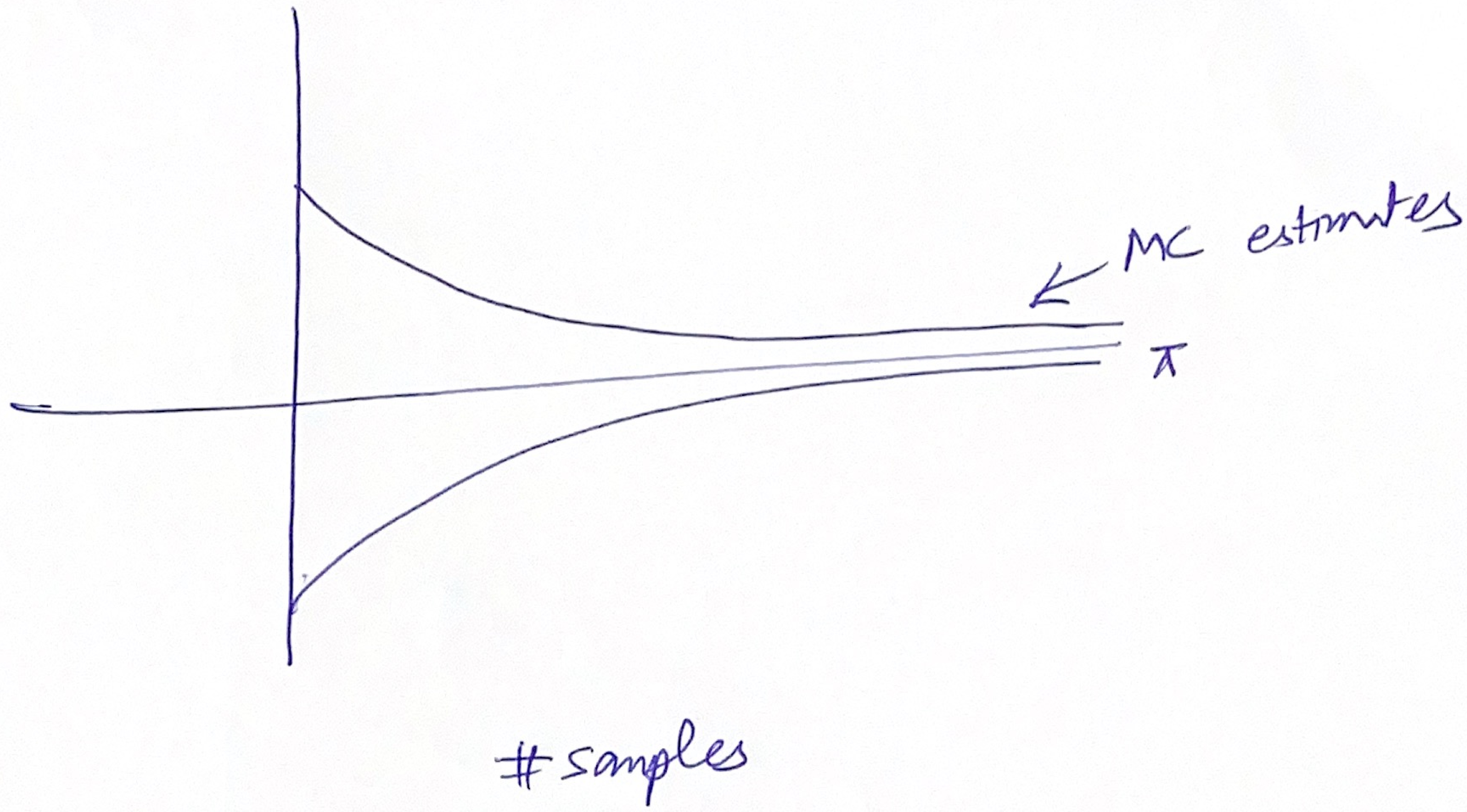
$$\hat{\phi} = \frac{\sum_{i=1}^{S} f(x_i)}{S}$$

$$E(\hat{\phi}) = \frac{1}{S} \int \sum_{i=1}^{S} f(x_i)\, p(x_i)\, dx_i$$

$$= \frac{1}{S} \sum_{i=1}^{S} \int f(x_i)\, p(x_i)\, dx_i$$

$$= \frac{1}{S} \sum_{i=1}^{S} E(f(x_i))$$

$$= \frac{1}{S} * S * \phi = \phi$$




$$\text{Variance} = E\left(\hat{\phi} - E(\hat{\phi})\right)^2$$

$$\propto \frac{1}{S}$$