1179: Probability
Lecture 26 — Hoeffding's Inequality and Weak Law of Large Numbers

Ping-Chun Hsieh (謝秉均)

December 17, 2021

Markov's inequality: X, non-negative

for any $a>0$:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Chebyshev's inequality: X, μ, σ^2

$$P(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \quad X_1, \ldots, X_n \text{ i.i.d.}$$

$$P(\underbrace{|\bar{X}-\mu| \geq t}) \leq \frac{\sigma^2}{t^2 \cdot n}$$

This Lecture

1. Hoeffding's Inequality

2. Weak Law of Large Numbers (WLLN)

- Reading material: Chapter 11.3-11.4

Review: Optimizing the Chernoff Bound

$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq e^{-ta} \cdot M_X(t)$, for any $t > 0$

- **Chernoff Bound**: Let X be a random variable with MGF $M_X(t)$. Suppose $M_X(t)$ exists for all t in some set S. Then, for any $t > 0$ and $t \in S$, for any $a \in \mathbb{R}$, we have

$$P(X \geq a) \leq e^{-\phi(a)},$$

where $\phi(a) = \max_{t>0, t \in S} (ta - \ln M_X(t))$

Example: Chernoff Bound for Bernoulli R.V.s

- Example: Suppose $X \sim \text{Bernoulli}(p)$
  - What is $M_X(t)$?
  - What is the Chernoff bound for X? $(P(X \geq a) \leq e^{-ta} \cdot M_X(t))$

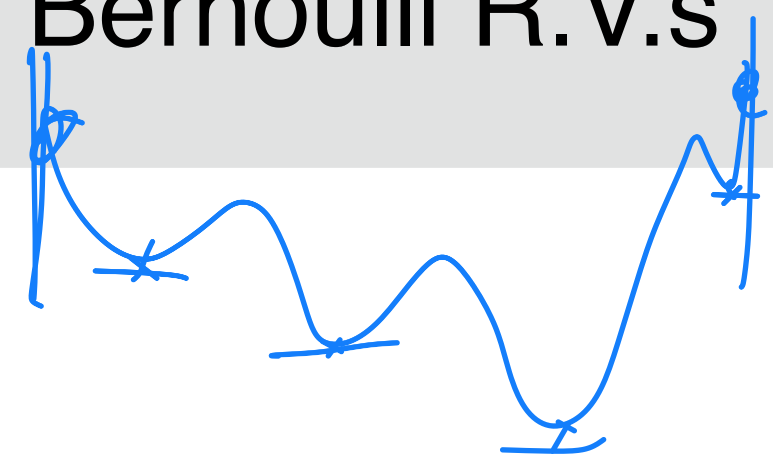
$$\begin{cases} X = 0, & \text{w.p. } 1-p \\ X = 1, & \text{w.p. } p \end{cases}$$

$$M_X(t) = E[e^{tX}] = (p)(e^{t \cdot 1}) + (1-p)(e^{t \cdot 0})$$

$$= p \cdot e^{t} + (1-p).$$

$$P(X \geq a) \leq e^{-ta} \cdot (p \cdot e^{t} + (1-p)) = p \cdot e^{t(1-a)} + (1-p) \cdot e^{-ta}$$

Example: Optimizing Chernoff Bound for Bernoulli R.V.s



- Example: Suppose $X \sim \text{Bernoulli}(p)$
  - How to optimize the Chernoff bound for X?
    $(P(X \geq a) \leq e^{-\phi(a)}, \phi(a) = \max_{t>0, t\in S} (ta - \ln M_X(t)))$

$$P(X \geq a) \leq p \cdot e^{t(1-a)} + (1-p) \cdot e^{-ta}$$

$$\Rightarrow \text{Minimize}_{t>0} \underbrace{p \cdot e^{t(1-a)} + (1-p) \cdot e^{-ta}}_{= g(t)}$$

Requirement of Chernoff bound

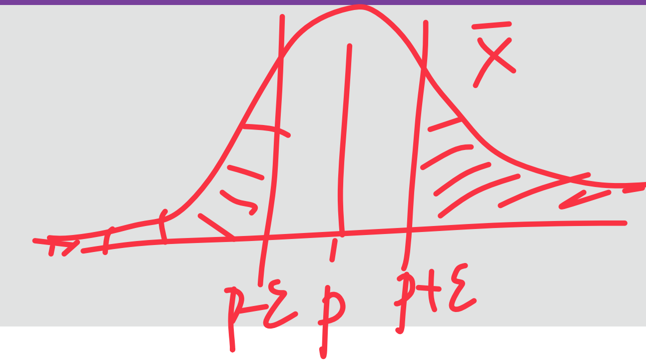
$$\frac{dg(t)}{dt} = (1-a) \cdot p\, e^{t(1-a)} - a(1-p)\, e^{-ta} = 0$$

$$\Rightarrow e^t = \frac{a(1-p)}{(1-a) \cdot p}$$

$$\Rightarrow \boxed{t = \ln \frac{a(1-p)}{(1-a) \cdot p}}$$

How about applying the Chernoff bound to “sum of independent random variables”?

Hoeffding's Inequality (Formally)



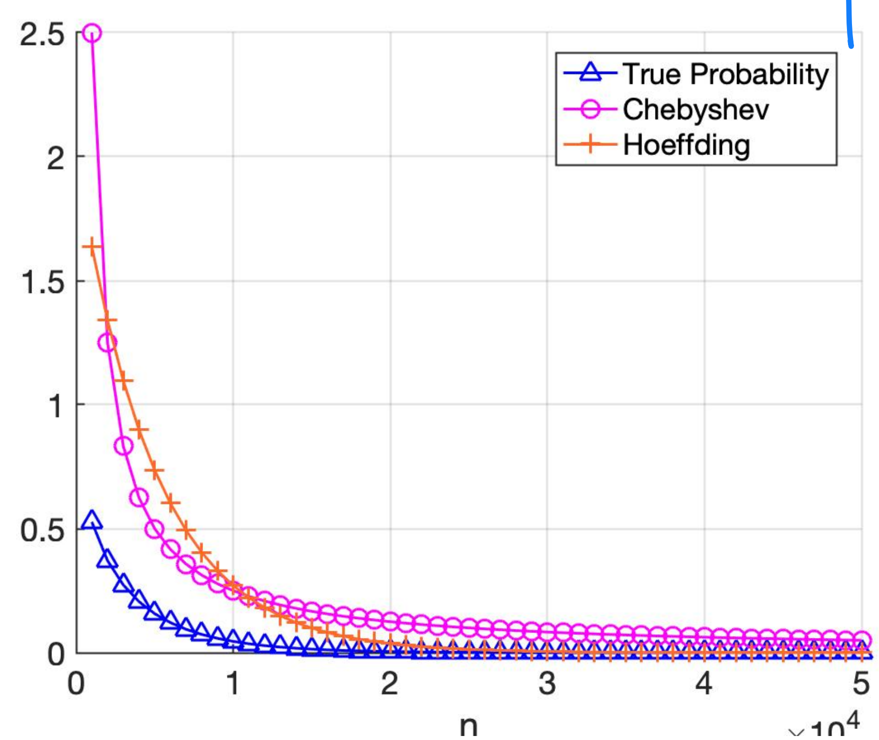
- **Hoeffding's Inequality (For Bernoulli)**: Let $X_1, \cdots, X_n$ be a sequence of i.i.d. Bernoulli random variables with parameter p. Define $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then, for any $\varepsilon > 0$, we have

ε small

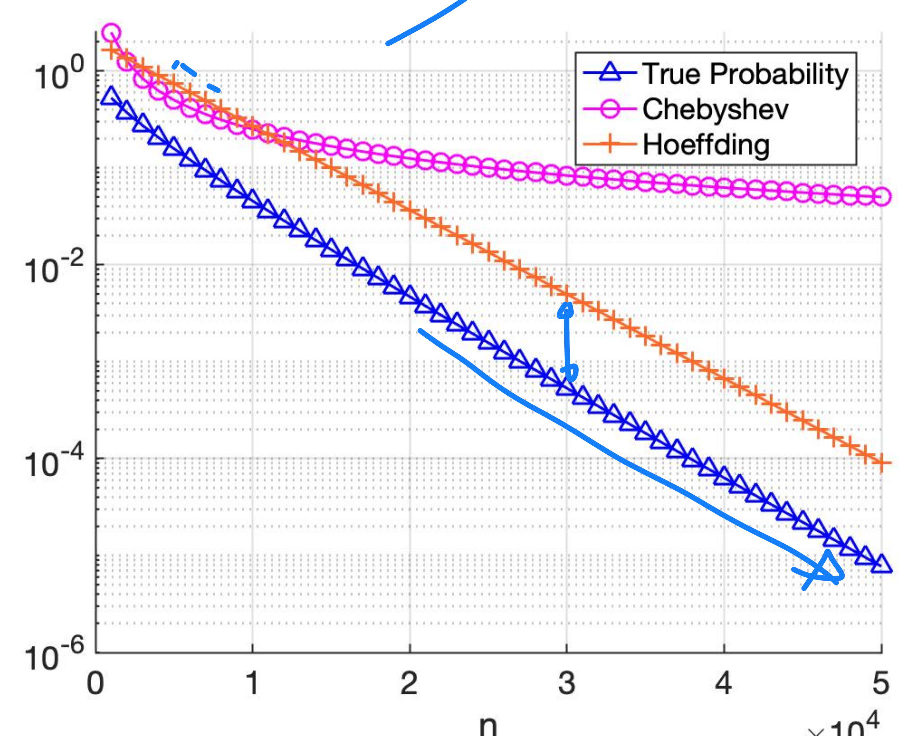
$$P(|\bar{X} - p| \geq \varepsilon) \leq 2\exp(-2n\varepsilon^2)$$

$P(\bar{X} - p \geq \varepsilon$ OR $\bar{X} - p \leq -\varepsilon)$

$\varepsilon = 0.01$



log scale



Proof of Hoeffding's Inequality (Positive Part)

$\bar{X} = \frac{1}{n}(X_1+X_2+\cdots+X_n)$

X, Y

$E[XY]$

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

$$P(\bar{X} - p \geq \varepsilon) \leq \exp(-2n\varepsilon^2)$$

- [Hint] Chernoff bound: $P(X \geq a) \leq e^{-ta} \cdot M_X(t)$

$P(\bar{X} - p \geq \varepsilon) \leq P(\underbrace{X_1+X_2+\cdots+X_n - np}_{=Z} \geq n\varepsilon)$

$\leq e^{-t\cdot n\varepsilon} \cdot M_Z(t) = e^{-t\cdot n\varepsilon} \cdot \left(E\left[e^{t(X_1-p)}\right]\right)^n$

Hoeffding's Lemma

$\leq e^{-tn\cdot\varepsilon} \cdot e^{\frac{nt^2}{8}}$

$= \boxed{e^{\frac{nt^2}{8} - tn\varepsilon}}$

$\frac{nt^2}{8} - tn\varepsilon = g(t)$

$g'(t) = \frac{n}{4}t - n\varepsilon = 0$

$t = 4\varepsilon$

$M_Z(t) = E\left[e^{t(X_1+X_2+\cdots+X_n-np)}\right]$

$= E\left[e^{t((X_1-p)+(X_2-p)+\cdots+(X_n-p))}\right]$

$= E\left[e^{t(X_1-p)}\right]\cdot E\left[e^{t(X_2-p)}\right]\cdots E\left[e^{t(X_n-p)}\right]$

$= \left(E\left[e^{t(X_1-p)}\right]\right)^n$

$$P(\overline{X}-p \geq \varepsilon) \leq e^{\boxed{\frac{nt^2}{8} - t \cdot n\varepsilon}}$$

$$\boxed{\frac{nt^2}{8} - t \cdot n\varepsilon} = g(t)$$

$$g'(t) = \frac{nt}{4} - n\varepsilon = 0$$

$$\Rightarrow \quad t = 4\varepsilon$$

$$P(\overline{X}-p \geq \varepsilon) \leq e^{\frac{n \cdot (4\varepsilon)^2}{8} - 4\varepsilon \cdot n \cdot \varepsilon}$$

$$= e^{-2 \cdot n \cdot \varepsilon^2}$$

Hoeffding's Lemma

- **Hoeffding's Lemma**: Let Z be a random variable with $E[Z] = 0$, and $Z \in [a, b]$ with probability 1. Then, for any $t > 0$, we have

$$E[e^{tZ}] \leq \exp\left(\frac{t^2(b-a)^2}{8}\right)$$

- **Question:** If $Z \sim \text{Bernoulli}(p)$, then $E[e^{t(Z-p)}] \leq \exp\left(\frac{t^2 \cdot (\ )^2}{8}\right) = e^{\frac{t^2}{8}}$

0 or 1

$-p$ $1-p$

Y

$Y = Z - p \Rightarrow E[Y] = 0$

$Y \in [\underbrace{-p}_{a}, \underbrace{1-p}_{b}]$

Proof of Hoeffding's Inequality (Negative Part)

$$P(\bar{X} - p \leq -\varepsilon) = P(p - \bar{X} \geq \varepsilon) \leq \exp(-2n\varepsilon^2)$$

- [Hint] Chernoff bound: $P(X \geq a) \leq e^{-ta} \cdot M_X(t)$

$P(p - \bar{X} \geq \varepsilon) \leq$

BJ4

Weak Law of Large Numbers (WLLN)

Review: Chebyshev's and Sample Mean: $n \to \infty$

- **Chebyshev's and Sample Mean**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. Define $S_n = (X_1 + \cdots + X_n)$. Then, for any $\varepsilon > 0$, we have

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon \right) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

- What if we let $n \to \infty$?

$$\lim_{n \to \infty} P\left(|\underbrace{\bar{X}}_{\text{empirical mean}} - \mu| \geq \varepsilon \right) = 0, \quad \text{for any } \varepsilon > 0$$

The Weak Law of Large Numbers (WLLN)

- **The Weak Law of Large Numbers (Khinchin's Law)**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, for every $\varepsilon > 0$, we have

$$P\left(\left| \frac{S_n}{n} - \mu \right| \geq \varepsilon \right) \to 0 \quad \text{as } n \to \infty$$

- Question: Any change in technical conditions (cf: Chebyshev's)?

No requirement on "finite variance".

- Question: What does "convergence" mean here?

Convergence in Probability

- **Convergence of a Deterministic Sequence**: Let $a_1, a_2 \cdots$ be a sequence of real numbers. We say that a_n converges to a if for every $\varepsilon > 0$, there exists N_0 such that

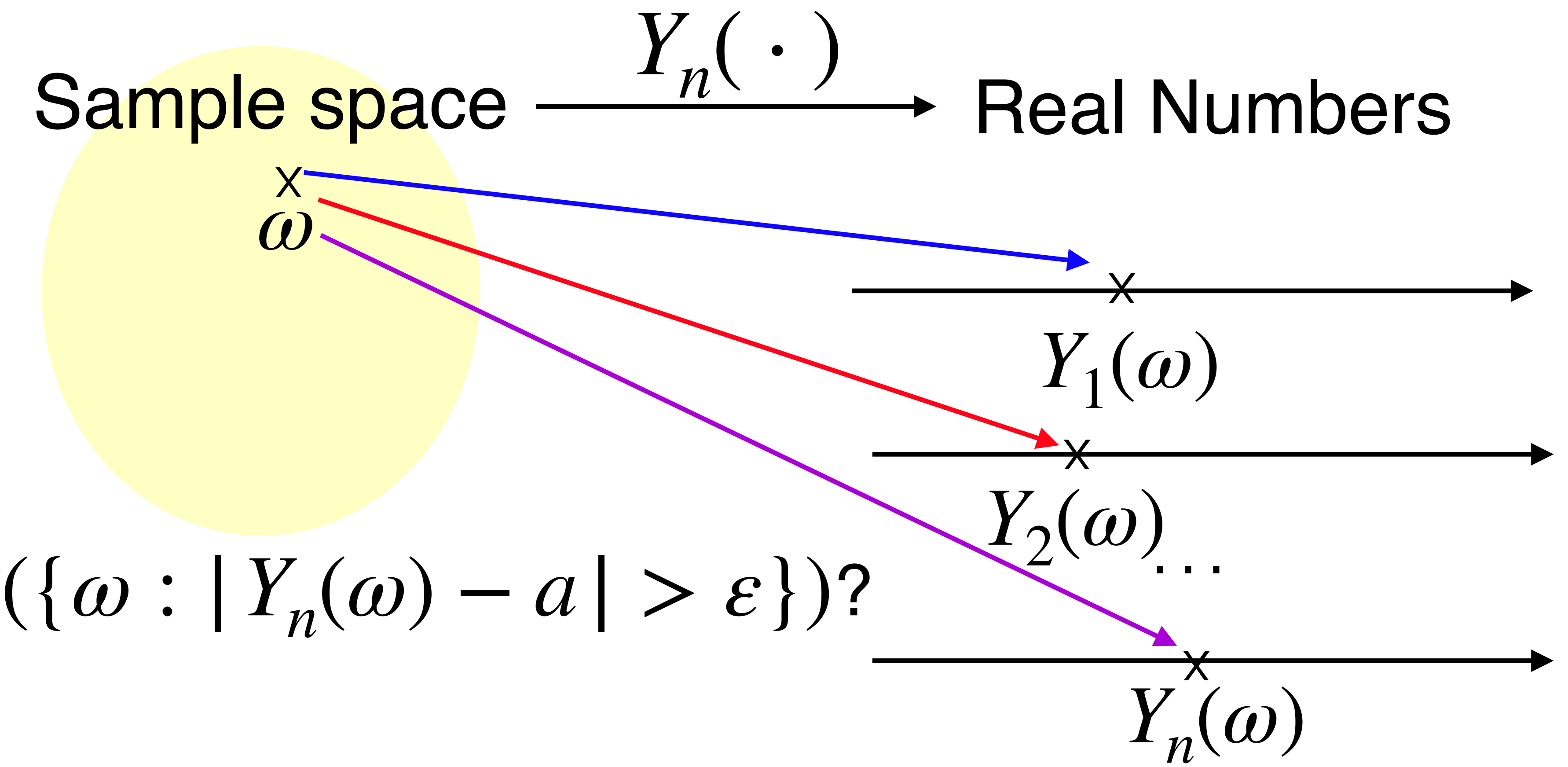
$$|a_n - a| \leq \varepsilon \qquad \text{for all } n \geq N_0$$

- **Convergence to a Scalar in Probability**: Let $Y_1, Y_2 \cdots$ be a sequence of random variables, and let a be a real number. We say that Y_n converges to a in probability if for every $\varepsilon > 0$,

- Question: How to interpret this definition?

Recall: Random Variables Defined on Ω

- $Y_1, Y_2, \cdots, Y_n, \cdots$ are defined on the same sample space Ω



- How to interpret $P(\{\omega : |Y_n(\omega) - a| > \varepsilon\})$?

- How about $\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - a| > \varepsilon\}) = 0$?

Example: Convergence in Probability

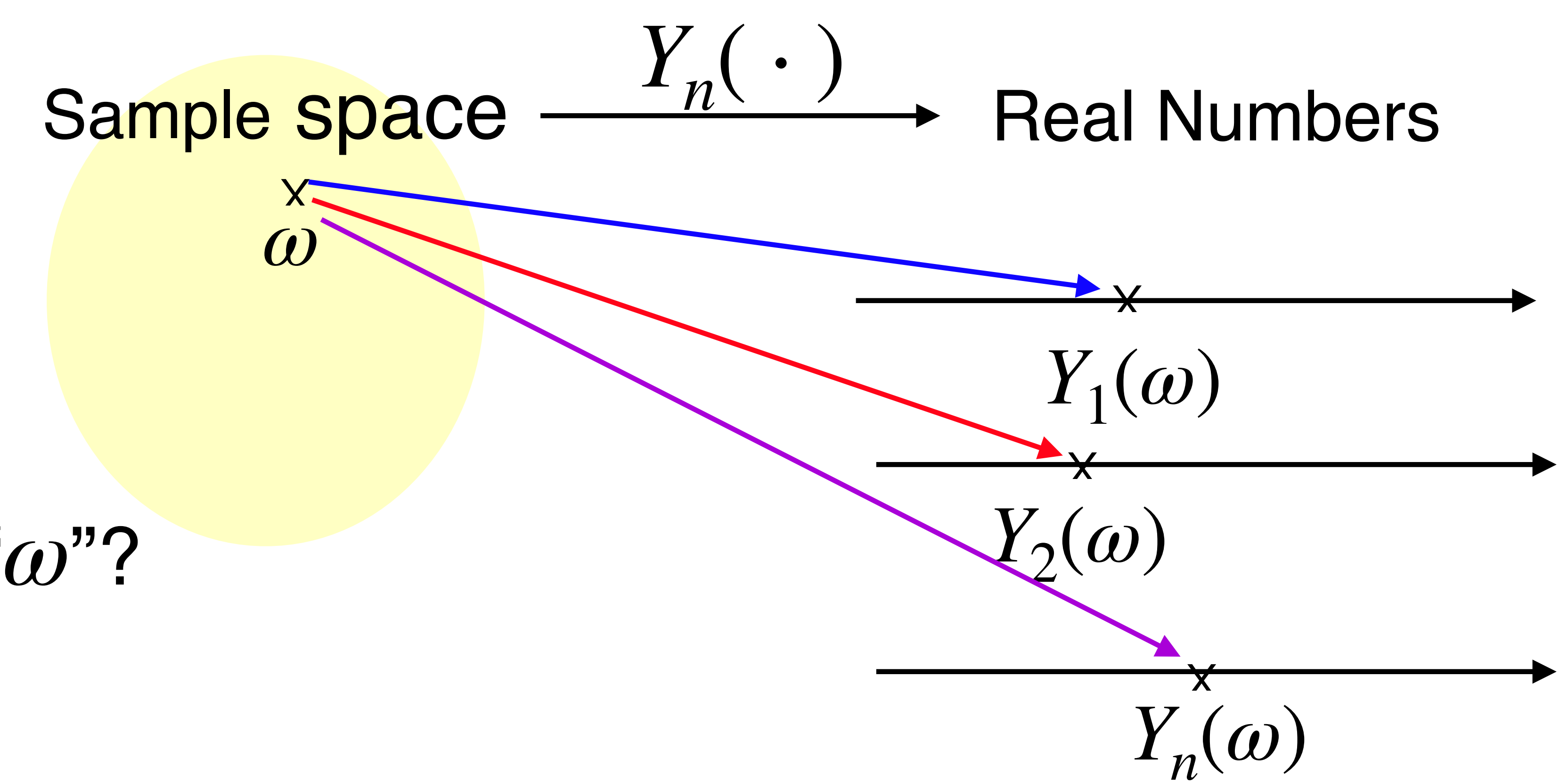
- Example: Consider a sequence of r.v.s Y_n

$$P(Y_n = y) = \begin{cases} 1 - \frac{1}{n} & \text{, if } y = 0 \\ \frac{1}{n} & \text{, if } y = n^2 \\ 0 & \text{, otherwise} \end{cases}$$

- For every $\varepsilon > 0$, can we find $P(|Y_n - 0| > \varepsilon)$?
- How about $\lim_{n \to \infty} P(|Y_n - 0| > \varepsilon)$?

How to Interpret WLLN?

- Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean μ
- Define $Y_n = (X_1 + X_2 \cdots, + X_n)/n$
- WLLN: $\lim_{n\to\infty} P(\{\omega : |Y_n(\omega) - \mu| > \varepsilon\}) = 0, \forall \varepsilon > 0$



- Question: What is an "ω"?

Rewriting WLLN (More Formally)

- **The Weak Law of Large Numbers (Khinchin's Law)**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. Define $S_n = (X_1 + \cdots + X_n)$. Then, for every $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P\Big(\{\omega : |\frac{S_n(\omega)}{n} - \mu| \geq \varepsilon\}\Big) = 0$$

In short, we have $\dfrac{S_n}{n} \xrightarrow{\text{p}} \mu$