1. Probability

▼ Inequality

- **Markov's Inequality:**

For any nonnegative random variable X and $t > 0$,

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

- **Chebyshev's Inequality:**

For any random variable X with finite second moment and $t > 0$.

$$\mathbb{P}[|X - E[X]| > t] \leq \frac{Var(X)}{t^2}$$

- **Chernoff Bound:**

X is random variable and for positive t this gives a bound on the upper tail of X in terms of its moment-generating function $M_X(t) = E[e^{tX}]$.

$$P(X \geq a) \leq inf_{t>0} \frac{M_X(t)}{e^{ta}} \qquad (t > 0)$$

- **Chernoff Bound for sum:**

Let $X_1, X_2, .., X_n$ be iid and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. For any $\lambda > 0$,

$$\begin{aligned}
\mathbb{P}\left[\bar{X}_n \geq t\right] &\leq \exp(-n\lambda t)\mathbb{E}\left[\exp\left(n\lambda\bar{X}_n\right)\right] && \text{Chernoff Inq.} \\
&= \exp(-n\lambda t)\mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\lambda X_i\right)\right] && \text{Definition of } \bar{X}_n \\
&= \exp(-n\lambda t)\prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\lambda X_i\right)\right] && \text{Independence of } X_i \\
&= \left[\exp(-\lambda t)\mathbb{E}\left[\exp\left(\lambda X_1\right)\right]\right]^n && \text{Identically distributed}
\end{aligned}$$

- **Fundamental Inequality:**

For any base b > 0 and x > 0,

$$(1 - \frac{1}{x})log_b(e) \leq log_b(x) \leq (x-1)log_b(e)$$

with equalities on both sides if, and only if, x = 1.

For natural log and x > 0,

$$(1 - \frac{1}{x}) \leq ln(x) \leq (x-1)$$

- **Gibbs' inequality:**

Suppose that $P = \{p_1, ..., p_n\}$ is a discrete probability distribution. Then for any other probability distribution $Q = \{q_1, ..., q_n\}$ he following inequality between positive quantities hold:

$$-\sum_{i=1}^{n} p_i log\ p_i \leq -\sum_{i=1}^{n} p_i log\ q_i$$

with equality if and only if $p_i = q_i$.

▼ Limit Theorems

- **Law of large numbers (LLN):**

The average of the results obtained from a large number of trials should be close to the expected value and tends to become closer to the expected value as more trials are performed.

$$lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i}{n} = \bar{X}$$

- Weak **Law of large numbers (WLLN):**

Let $X_1, X_2, .., X_n$ be i.i.d random variables with a finite expected value $EX_i = \mu < \infty$. Then for any $\epsilon > 0$.

The WLLN states that the average of a large number of i.i.d. random variables converges in probability to the expected value.

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

- **Central limit theorem (CLT):**

T**he sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough**. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

The CLT states that <u>the normalized average of a sequence of i.i.d. random variables converges in distribution to a standard normal distribution.</u>

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

▼ Convergence of Random Variables

In some situations, we would like to see if a sequence of random variables $X_1, X_2, X_3, ...$ "converges" to a random variable X. That is, we would like to see if X_n gets closer and closer to X in some sense as n increases.

- **Convergence of a Sequence of Numbers**

A sequence $a_1, a_2, a_3, ...$ converges to a limit L if

$$\lim_{n\to\infty} a_n = L$$

That is, for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that

$$|a_n - L| < \epsilon, \quad \text{for all } n > N$$

- **Sequence of Random Variables**

We have sample space S and probability measure P .

A sequence of random variables is in fact a sequence of functions $X_n : S \to \mathbb{R}$.

- **Convergence in Distribution**

A sequence of random variables $X_1, X_2, X_3, ...$ converges in distribution to a random variable X, shown by $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for all x at which $F_X(x)$ is continuous.

- **Convergence in Probability**

A sequence of random variables $X_1, X_2, X_3, ...$ converges in probability to a random variable X , shown by $X_n \xrightarrow{p} X$, if

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0, \quad \text{for all } \epsilon > 0$$

- **Convergence in Mean**

Let $r > 1$ be a fixed number. A sequence of random variables $X_1, X_2, X_3, ...$ converges in the r^{th} mean or in the L^r norm to a random variable X, shown by $X_n \xrightarrow{L^r} X$, if

$$\lim_{n\to\infty} E(|X_n - X|^r) = 0$$

if $r = 2$, it is called the **mean-square convergence**, and it is shown by $X_n \xrightarrow{m.s.} X$.

- Almost Sure Convergence

A sequence of random variables $X_1, X_2, X_3, ...$ converges **almost surely** to a random variable X, shown by $X_n \xrightarrow{a.s.} X$, if

$$P(\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\}) = 1$$

▼ The Typical Set and AEP

Let $X_1, X_2, X_3, ...$ be iid copies of a random variable $X \sim p(x)$ with finite support $\mathcal{X}$.

- **High Probability Sets**

1. A length-n random sequence is denote by $X_n = (X_1, X_2, ..., X_n)$ and a realization is denote by $x^n = (x_1, x_2, ..., x_n)$, the joint pmf is the product measure $p_{X^n}()x^n = \mathbb{P}[X^n = x^n] = \prod_{i=1}^{n} p(x_i) = p(x^n)$.
2. Our goal is to identify a subset of $A \in \mathcal{X}^n$ which contains most of the probability.

3. It is useful to define such a set in terms of a function $g : \mathbb{R} \to \mathbb{R}$. By the law of large numbers, for any $\epsilon > 0$ there exists N such that:

$$\mathbb{P}[|\frac{1}{n}\sum_{i=1}^{n} g(x_i) - \mathbb{E}[g(X)]| \leq \epsilon] \geq 1 - \epsilon, \quad \text{for any } n \geq N_\epsilon$$

4. This means that almost all of the probability is concentrated on the set of sequences $A \in \mathcal{X}^n$ given by

$$A = \{x^n \in \mathcal{X}^n : |\frac{1}{n}\sum_{i=1}^{n} g(x_i) - \mathbb{E}[g(X)]| \leq \epsilon\}$$

   Equivalently

$$\mathbb{E}[g(X)] - \epsilon \leq \frac{1}{n}\sum_{i=1}^{n} g(x_i) \leq \mathbb{E}[g(X)] + \epsilon$$

   Or

$$2^{-n(\mathbb{E}[g(X)]+\epsilon)} \leq 2^{-\sum_{i=1}^{n} g(x_i)} \leq 2^{-n(\mathbb{E}[g(X)]-\epsilon)}$$

- The Typical Set

For the special choice of $g(x) = -log\ p(x)$, it follow that:

$$2^{-\sum_{i=1}^{n} g(x_i)} = 2^{\sum_{i=1}^{n} log\ p(x_i)} = \prod_{i=1}^{n} p(x_i) = p_{X_n}(x^n)$$

and

$$\mathbb{E}[g(X)] = \mathbb{E}[-log\ p(X)] = \underbrace{H(X)}_{\text{entropy of p(x)}}$$

Definition: The ϵ-typical set is defined by

$$A_\epsilon^{(n)} = \{x^n \in \mathcal{X}^n : 2^{-n(H(X)+\epsilon)} \leq p_{X^n}(x^n) \leq 2^{-n(H(X)-\epsilon)}\}$$

This is the set of all sequences whose probability is approximately equal to the expected probability. Unusually likely and unusually unlikely sequences are excluded.

💡 **The typical set contains almost all of the probability. Furthermore, all of the sequences in the typical set have roughly the same probability. This is know as the Asymptotic Equipartition property (AEP)**

By Law of Large Numbers:

$$\mathbb{P}[X^n \in A_\epsilon^{(n)}] \geq 1 - \epsilon \ \text{ for all } n \geq N_\epsilon$$

Upper Bound:

$$|A_\epsilon^{(n)}| \leq 2^{n(H(x)+\epsilon)}$$

▼ Proof

$$\begin{aligned} 1 &= \sum_{x \in \mathcal{X}^n} p(x^n) \\ &\geq \sum_{x \in A_\epsilon^n} p(x^n) \\ &\geq \sum_{x \in A_\epsilon^n} 2^{-n(H(x)+\epsilon)} \\ &= |A_\epsilon^n| \cdot 2^{-n(H(x)+\epsilon)} \end{aligned}$$

$$\begin{aligned} |A_\epsilon^n| \cdot 2^{-n(H(x)+\epsilon)} &\leq 1 \\ |A_\epsilon^n| &\leq 2^{n(H(x)+\epsilon)} \end{aligned}$$

Lower Bound:

$$|A_\epsilon^{(n)}| \geq (1-\epsilon)2^{n(H(x)-\epsilon)}$$

▼ Proof

$$\begin{aligned} 1-\epsilon &\leq \mathbb{P}[X^n \in A_\epsilon^{(n)}] \\ &= \sum_{x \in A_\epsilon^n} p(x^n) \\ &\leq \sum_{x \in A_\epsilon^n} 2^{-n(H(X)-\epsilon)} \\ &= |A_\epsilon^n| \cdot 2^{-n(H(X)-\epsilon)} \end{aligned}$$

$$\begin{aligned} 1-\epsilon &\leq |A_\epsilon^n| \cdot 2^{-n(H(X)-\epsilon)} \\ (1-\epsilon)2^{n(H(X)-\epsilon)} &\leq |A_\epsilon^n| \end{aligned}$$

Cesaro Means:

Let $a_n \to a$, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} a_i = a$$

Bernoulli distribution:

pmf:

$$f(k;p) = \begin{cases} p, & \text{if } \text{k=1} \\ q = 1-p, & \text{if } \text{k=0} \end{cases}$$