Jensen **不等式**

Jensen **不等式**是凸函数（或凹函数）的一个基本性质，在概率论和优化问题中有重要应用。它给出了凸函数作用于随机变量的期望的一个基本上界或下界。

---

1. Jensen **不等式的数学表达**

设:

- f 是一个定义在区间 $I \subseteq \mathbb{R}$ 上的**凸函数**;
- X 是一个取值在 I 上的随机变量;
- $\mathbb{E}[X]$ 存在（有限值）。

  则:

  $$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

  如果 f 是凹函数，则不等式方向相反:

  $$f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)].$$

---

2. **凸函数和凹函数的定义**

- 函数 $f(x)$ 是凸的，当且仅当对任意 $x_1, x_2 \in I$ 和 $\theta \in [0, 1]$，有:

  $$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta)f(x_2).$$

  直观上，凸函数的图像总是**在连线的下方**。

- 函数 $f(x)$ 是凹的，当且仅当 $-f(x)$ 是凸函数。

---

3. Jensen **不等式的直观解释**

Jensen 不等式告诉我们：

- 对一个**凸函数** f 来说，先取期望再计算函数值（即 $f(\mathbb{E}[X])$）的结果小于或等于先计算函数值再取期望（即 $\mathbb{E}[f(X)]$）。
- 这种性质源于凸函数的"向上弯曲"形状——它更容易被"拉高"，所以 $\mathbb{E}[f(X)]$ 通常更大。

例如：

- 假设 X 是一个随机变量, 服从标准正态分布 (均值为 0), 而 $f(x) = x^2$ 是一个凸函数。我们可以验证 Jensen 不等式：

$$f(\mathbb{E}[X]) = f(0) = 0,$$

$$\mathbb{E}[f(X)] = \mathbb{E}[X^2] = \mathrm{Var}(X) = 1.$$

确实有 $0 \leq 1$。

---

4. Jensen **不等式的证明**

基本证明（离散情形）

设 X 是离散随机变量，取值 $x_1, x_2, \ldots, x_n$ 的概率分别为 $p_1, p_2, \ldots, p_n$。则 $\mathbb{E}[X] = \sum_{i=1}^{n} p_i x_i$，而我们需要证明：

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i).$$

由凸函数的定义：

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i),$$

其中使用了 f 对凸组合的单调性，这就是 Jensen 不等式。

基本证明（连续情形）

如果 X 是连续随机变量，取值在区间 I 上，其概率密度函数为 $p(x)$，则：

$$\mathbb{E}[X] = \int_I xp(x)\,dx, \quad \mathbb{E}[f(X)] = \int_I f(x)p(x)\,dx.$$

同样由凸函数的定义以及积分线性性：

$$f\left(\int_I xp(x)\,dx\right) \leq \int_I f(x)p(x)\,dx.$$

这就是连续情况下的 Jensen 不等式。

---

5. Jensen **不等式的应用**

(1) **信息论**

在信息论中，熵（Entropy）的定义用到了 Jensen 不等式。例如，对于一个随机变量 X，其熵为：

$$H(X) = -\sum_x p(x)\log p(x).$$

使用 Jensen 不等式可以证明熵的非负性。

(2) Hoeffding **引理**

Hoeffding 引理中，使用 Jensen 不等式来上界 $\mathbb{E}[e^{\lambda Z}]$。通过凸函数的性质，可以将 Z 的有界性引入到期望上界中。

(3) **凸优化**

在凸优化中，Jensen 不等式用来证明凸目标函数的许多性质，例如：凸函数的组合仍然是凸的。

---

6. **示例**

示例 1：平方函数

设 X 是一个随机变量，$f(x) = x^2$ 是一个凸函数。则 Jensen 不等式表明：

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

展开后是：

$$(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2].$$

这实际上是**方差非负性**的本质：因为

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0.$$

示例 2：对数函数

设 $f(x) = \log(x)$，这是一个凹函数。则 Jensen 不等式表明：

$$\log(\mathbb{E}[X]) \geq \mathbb{E}[\log(X)].$$

这在信息论中广泛应用，例如处理加权平均的对数上界问题。

---

7. 总结

Jensen 不等式本质上是凸函数（或凹函数）在概率期望中的基本性质，广泛应用于信息论、凸优化、统计学习理论等领域。