第 2 讲 期望与方差, 概率极限定理

设 X 有离散的概率分布

$$p_j = P(X = x_j),\ \ j = 0, 1, \cdots .$$

如果 $P(X \geq 0) = 1$ 或级数 $\sum_{j=0}^{\infty} |x_j| p_j$ 收敛，则称 $E[X] = \sum_{j=0}^{\infty} x_j p_j$ 为 X 的**数学期望或均值**.

设 X 是有密度函数 $f(x)$ 的随机变量. 如果 $P(X \geq 0) = 1$ 或 $\int_{-\infty}^{\infty} |x| df(x) < \infty$, 则称 $E[X] = \int_{-\infty}^{\infty} x f(x) dx$ 为 X 的**数学期望或均值**.

设 $f_0(s)$ 是非负函数，如果随机变量 X 的分布函数 $F(x)$ 可以分解成

$$F(x) = F_1(x) + F_2(x),\ x \in (-\infty, \infty),$$

其中 $F_1(x) = \int_{-\infty}^{x} f_0(s)ds$， $F_2(x)$ 是仅在至多可数多个点 x_j 处有跳跃高度 $P(X_j = x_j) = F(x_j) - F(x_j-)$ 的阶梯函数，则称 X 有**混合分布**. 如果 X 非负或

$$\int_{-\infty}^{\infty} |x| f_0(x)dx + \sum_{j\geq 1} |x_j| P(X = x_j) < \infty,$$

则称 X 的数学期望存在，并且定义 X 的数学期望为

$$E[X] = \int_{-\infty}^{\infty} x f_0(x)dx + \sum_{j\geq 1} x_j P(X = x_j).$$

对于任何右连续的单调不减阶梯函数 $G(x)$，设 G 仅在至多可数多个点 x_j 处有跳跃. 对于另一函数 $g(x)$，如果

$$\sum_{j:x_j\in[a,b]} |g(x_j)|[G(x_j)-G(x_j-)] < \infty,\ \ a<b, \tag{1.1}$$

则称 $g(x)$ 在 $[a,b]$ 上关于 G 可积，并定义积分

$$\int_a^b g(x)dG(x) = \sum_{j:x_j\in[a,b]} g(x_j)[G(x_j)-G(x_j-)].$$

按照上面的积分定义，如果 X 的数学期望存在，则可以把 X 的数学期望表示成

$$E[X] := \int_{-\infty}^{\infty} xdF_1(x) + \int_{-\infty}^{\infty} xdF_2(x) := \int_{-\infty}^{\infty} xdF(x).$$

积分 (1.1) 有如下的基本性质:

(1) 若 $g(t), h(t)$ 都在 $[a, b]$ 上关于 G 可积，则有

$$\int_a^b (g(t)+h(t))dG(t) = \int_a^b g(t)dG(t) + \int_a^b h(t)dG(t);$$

(2) 若 $g(t)$ 在 $[a, b]$ 上关于 G, F 可积，则有

$$\int_a^b g(t)d[G(t)+F(t)] = \int_a^b g(t)dG(t) + \int_a^b g(t)dF(t);$$

(3) 设 $F_n(t) = P(X_n \leq t)$，若 $m(t) = \sum_{n=1}^{\infty} F_n(t) < \infty$ 对 $t < \infty$ 成立，则对于非负函数 $g(t)$，有

$$\sum_{n=1}^{\infty} \int_{-\infty}^{\infty} g(t)dF_n(t) = \int_{-\infty}^{\infty} g(t)dm(t). \tag{1.2}$$

设 X 是一个随机变量，X 的方差定义为

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

两个定义在相同样本空间上的随机变量 X 和 Y 称为不相关的，如果他们的协方差

$$\mathrm{Cov}(X,Y)=E[(X-E[X])(Y-E[Y])]=E[XY]-E[X]E[Y]$$

等于 0. 由定义可知独立的随机变量是不相关的. 然而，不相关的随机变量不一定是独立的.

命题 1.1

如果 $X_1, X_2, \cdots, X_n$ 都是定义在同一样本空间的随机变量，则有

$$\mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{j=1}^{n}\sum_{i=1}^{j-1} \mathrm{Cov}(X_i, X_j).$$

例 1.2

设 X 服从两点分布：$X \sim \mathcal{B}(1, p)$. 我们有 $E[X] = p$，$\mathrm{Var}(X) = p(1-p)$.

例 1.3

如果 $Y_1, Y_2, \cdots, Y_n$ 是来自总体 $\mathcal{B}(1, p)$ 的样本，则 $\xi_n = Y_1 + Y_2 + \cdots + Y_n \sim \mathcal{B}(n, p)$. 我们有 $E[\xi_n] = np$，$\mathrm{Var}(\xi_n) = npq$.

例 1.4

设随机变量 X 服从几何分布：$X \sim \text{Geom}(p)$. 我们有 $E[X] = 1/p$，$\text{Var}(X) = q/p^2$.

例 1.5

设随机变量 X 服从泊松分布：$X \sim \mathcal{P}(\lambda)$，则 $E[X] = \lambda$，$\text{Var}[X] = \lambda$.

例 1.6

设 $X \sim \mathcal{E}(\lambda)$，则 $E[X] = 1/\lambda$，$\text{Var}(X) = 1/\lambda^2$.

设 $\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_n)$ 是随机向量，A 是随机事件. 如果 $g(\boldsymbol{y})=P(A|\boldsymbol{Y}=\boldsymbol{y})$，则定义

$$P(A|\boldsymbol{Y})=g(\boldsymbol{Y}),$$

并且称 $P(A|\boldsymbol{Y})$ 是已知 $\boldsymbol{Y}$ 时 A 的条件概率，简称为条件概率.

命题 2.1

条件概率有如下的基本性质:

$$E[P(A|\boldsymbol{Y})] = P(A).$$

设 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)$ 是随机向量，X 是随机变量，满足 $E[|X|]<\infty$. 如果 $g(\boldsymbol{y})=E[X|\mathbf{Y}=\boldsymbol{y}]$，我们定义

$$E[X|\boldsymbol{Y}]=g(\boldsymbol{Y}),$$

并且称 $E[X|\boldsymbol{Y}]$ 是已知 $\boldsymbol{Y}$ 时 X 的条件数学期望. 条件数学期望有如下的基本性质: 设 $\boldsymbol{Y}=(Y_1,\cdots,Y_n)$ 是随机向量，X,Z 是随机变量，满足 $E[|X|]<\infty$，$E[|Z|]<\infty$.

设 $h(y_1, y_2, \cdots, y_n)$ 是一个实函数. 我们有

(1) $E[E[X|\boldsymbol{Y}]] = E[X]$;

(2) 当 X 与 $\boldsymbol{Y}$ 独立时，有 $E[X|\boldsymbol{Y}] = E[X]$;

(3) 当 Z 与 $(X, \boldsymbol{Y})$ 独立时，有 $E[XZ|\boldsymbol{Y}] = E[Z]E[X|\boldsymbol{Y}]$;

(4) 当 $Z = h(\boldsymbol{Y})$ 时，有 $E[XZ|\boldsymbol{Y}] = ZE[X|\boldsymbol{Y}]$;

(5) 对于常数 a, b，有 $E[aX + bZ|\boldsymbol{Y}] = aE[X|\boldsymbol{Y}] + bE[Z|\boldsymbol{Y}]$.

设 X 是一个随机变量，有分布函数 $F(x)$ 和生存函数 $\overline{F}(x) = 1 - F(x)$.

定理 3.1

当 X 只取非负整数值时，有

$$E[X] = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k).$$

定理 3.2

如果 $P(X \geq 0) = 1$，则有

$$E[X] = \int_0^\infty P(X > x)dx = \int_0^\infty \overline{F}(x)dx,$$

$$E[X^\alpha] = \int_0^\infty \alpha x^{\alpha-1}\overline{F}(x)dx, \ \alpha > 0.$$

定理 3.3

如果 $E[|g(X)|] < \infty$，$P(A) > 0$，则

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)dF(x),\ E[g(X)|A] = \frac{g(X)I[A]}{P(A)}.$$

其中 $I[A]$ 是事件 A 的示性函数，也可用 I_A 表示，即

$$I_A = I[A] = \begin{cases} 1, & \text{当 } A \text{ 发生}, \\ 0, & \text{当 } A \text{ 不发生}. \end{cases}$$

若 X 具有密度函数 $f(x)$，则

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

定理 3.4

对于随机事件 B，有全概率公式 $P(B) = \int_{-\infty}^{\infty} P(B|X = x)dF(x)$.

定理 3.5

若随机变量 Y 的数学期望 $E[Y]$ 存在，则有全概率公式

$$E[Y] = \int_{-\infty}^{\infty} E[Y|X = x]dF(x).$$

对于有联合分布 $F(\boldsymbol{x})$ 的随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$，有全概率公式

$$E[Y] = \int_{\mathbb{R}^n} E[Y|\boldsymbol{X} = \boldsymbol{x}]d\boldsymbol{F}(\boldsymbol{x}).$$

特别当 $\boldsymbol{X}$ 有联合密度 $f(\boldsymbol{x})$ 时，有

$$E[Y] = \int_{\mathbb{R}^n} E[Y|\boldsymbol{X} = \boldsymbol{x}]f(\boldsymbol{x})dx_1dx_2\cdots dx_n.$$

定理 3.6

对于随机变量 T，已知 $T=t$ 的条件下，$P_t(A)=P(A|T=t)$ 是概率. 对于任何事件 A,B，有

$$P_t(B|A)=P(B|A,T=t).$$

概率 $P_t(A)$ 只是表示已知 $T=t$ 的条件下事件 A 的发生概率, 在我们遇到的情况下都有明确的意义. 所以无论 $P(T=t)=0$ 与否, 我们都视 $P_t(\cdot)$ 为概率.

定理 3.7

当 Y 的数学期望 $E[Y]$ 存在时，用 $E[Y|T=t]$ 表示已知 $T=t$ 时 Y 的条件数学期望，则对于任意 n 维向量 $\boldsymbol{X}$ 及其在条件 $T=t$ 下的分布函数 $F_t(\boldsymbol{x})$，有

$$E[Y|T=t]=\int_{\mathbb{R}^n} E[Y|T=t,\boldsymbol{X}=\boldsymbol{x}]dF_t(\boldsymbol{x}).$$

证明.

我们有

$$\begin{aligned}
E[Y|T=t] &= \int_0^\infty P_t(Y>y)dy \\
&= \int_0^\infty \int_{\mathbb{R}^n} P_t(Y>y|\boldsymbol{X}=\boldsymbol{x})dF_t(\boldsymbol{x})dy \\
&= \int_{\mathbb{R}^n} (\int_0^\infty (Y>y|T=t,\boldsymbol{X}=\boldsymbol{x})dy)dF_t(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^n} E(Y|T=t,\boldsymbol{X}=\boldsymbol{x})dF_t(\boldsymbol{x}).
\end{aligned}$$

从而得证. □

特别地, 若 X 是离散随机变量, 我们有:

定理 3.8

当 Y 的数学期望 $E[Y]$ 存在时, 用 $E[Y|T=t]$ 表示已知 $T=t$ 时 Y 的条件数学期望, 则对于任意离散型随机变量 X 及其概率分布 $p_j=P(X=x_j)$, $j=1,2,\cdots$, 有

$$E[Y=t|T=t]=\sum_{j=1}^{\infty}E[Y|T=t,X=x_j]P(X=x_j|T=t).$$

1. 随机变量的数学期望和方差
2. 条件概率和条件数学期望
3. 数学期望的计算公式
4. 特征函数与概率极限定理
   - 概率不等式
   - 特征函数
   - 概率极限定理
5. 次序统计量

定理 4.1: 马尔可夫不等式

对随机变量 X 和常数 $\epsilon > 0$，有

$$P(|X| \geq \epsilon) \leq \frac{1}{\epsilon^{\alpha}} E[|X|^{\alpha}],\ \ \alpha > 0.$$

推论 4.2: 切比雪夫不等式

对随机变量 X 和常数 $\epsilon > 0$，有

$$P(|X - E[X]| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathrm{Var}(X).$$

证明.

在马尔可夫不等式中取 $\alpha = 2$ 并用 $X - E[X]$ 代替 X 可得. □

定理 4.3

设 $E[X^2]<\infty$，$E[Y^2]<\infty$，则有

$$|E[XY]| \le \sqrt{E[X^2]E[Y^2]}.$$

并且上面不等式中的等号成立的充分必要条件是有不全为零的常数 a, b，使得 $aX+bY=0$ a.s..

定理 4.4

若 f 是凸函数，则只要期望存在，就有

$$E[f(X)] \geq f(E[X]).$$

如果 ξ, η 是随机变量，$i=\sqrt{-1}$，则称

$$Z=\xi+i\eta$$

是复复值随机变量. 如果 $E[\xi]$，$E[\eta]$ 存在，则定义 Z 的数学期望为

$$E[Z]=E[\xi]+iE[\eta].$$

对随机变量 X，因为 $\sin(tX)$，$\cos(tX)$ 的数学期望都存在，所以定义

$$\phi(t)=E[e^{itX}]=E[\cos(tX)]+iE[\sin(tX)],\ t\in\mathbb{R}.$$

上面定义的函数 $\phi(t)$ 称为 X 的特征函数. 可以证明随机变量的特征函数可以唯一决定该随机变量的分布函数. 所以, **随机变量的特征函数和分布函数相互唯一决定**.

例 4.5

用 $\phi(t)$ 表示 X 的特征函数，通过计算可以得到：

(1) 如果 $X \sim \mathcal{B}(n,p)$，那么

$$\phi(t) = (pe^{it} + q)^n;$$

(2) 如果 $X \sim \text{Geom}(n,p)$，那么

$$\phi(t) = \frac{pe^{it}}{1 - qe^{it}};$$

(3) 如果 $X \sim \mathcal{P}(\lambda)$，那么

$$\phi(t) = \exp\{\lambda(e^{it} - 1)\};$$

(4) 如果 $X \sim U[a,b]$，那么

$$\phi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)};$$

(5) 如果 $X \sim \mathcal{E}(\lambda)$，则

$$\phi(t) = (1 - it/\lambda)^{-1};$$

(6) 如果 $X \sim N(\mu, \sigma^2)$，则

$$\phi(t) = \exp(i\mu t - \sigma^2 t^2/2).$$

命题 4.6

用 $\phi(t)$ 表示 X 的特征函数，如果 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的随机变量，则 $Y = X_1 + X_2 + \cdots + X_n$ 有特征函数

$$\phi_Y(t) = [\phi(t)]^n.$$

命题 4.7

如果 X_i 有特征函数 $\phi_i(t)$,$X_1, X_2, \cdots, X_n$ 相互独立,则 $Y = X_1+X_2+\cdots+X_n$ 有特征函数

$$\phi_Y(t) = \phi_1(t)\phi_2(t)\cdots\phi_n(t).$$

我们也可以定义随机变量 $X_1, \cdots, X_n$ 的联合特征函数为

$$\phi(t_1, \cdots, t_n) = E\left[\exp\{i \sum_{j=1}^{n} t_j X_j\}\right]$$

可以证明联合特征函数唯一地确定联合分布.

设 $\{X_n\}$ 是随机变量序列，X 是随机变量或常数. 如果

$$P(\lim_{n\to\infty} X_n = X) = 1,$$

则称 X_n **几乎处处收敛**到 X 或**依概率 1 收敛**到 X，记做 $X_n \to X$ a.s.

设 $\{X_n\}$ 是随机变量序列，X 是随机变量或常数. 如果

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

则称随机变量序列 X_n **依概率收敛于** X，记为 $X_n \xrightarrow{P} X$.

对于分布函数列 $\{F_n(x)\}$，如果存在一个不减函数 $F(x)$ 使得

$$\lim_{n\to\infty} F_n(x) = F(x)$$

在 $F(x)$ 的每一个连续点上都成立，则称 $F_n(x)$ **弱收敛于** $F(x)$，并记为 $F_n(x) \xrightarrow{w} F(x)$. 设随机变量 X_n，X 的分布函数分别为 $F_n(x)$ 及 $F(x)$，如果 $F_n(x) \xrightarrow{w} F(x)$，则称 X_n **依分布收敛于** X，记为 $X_n \xrightarrow{d} X$.

定理 4.8: 强大数律

如果 $\{X_j\}$ 是独立同分布的随机变量序列，$\mu = EX_1$，则

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \ a.s.$$

定理 4.9: 中心极限定理

设随机变量序列 $\{X_j\}$ 独立同分布，有共同的数学期望 μ 和有限方差 σ^2，样本均值 $\overline{X}_n$ 和样本方差 $\hat{\sigma}^2$ 分别定义为

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j,$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2,$$

则 $\xi_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$，$\eta_n = \frac{\overline{X}_n - \mu}{\hat{\sigma}/\sqrt{n}}$ 都依分布收敛到标准正态分布，即对任意 x，当 $x_n \to x$ 时，

$$\lim_{n\to\infty} P(\xi_n \le x_n) = \Phi(x), \quad \lim_{n\to\infty} P(\eta_n \le x_n) = \Phi(x),$$

这里 $\Phi(x)$ 是标准正态分布的分布函数.

定理 4.10: 单调收敛定理

设 $\{X_n\}$ 是随机变量序列，满足 $0 \le X_1 \le X_2 \le \cdots$，$X$ 是随机变量，并且 $X_n \to X$ a.s.，则

$$EX_n \to EX.$$

对非负随机变量 $X_1, X_2, \cdots$，有

$$E[\sum_{j=1}^{\infty} X_j] = \sum_{j=1}^{\infty} E[X_j].$$

对于非负函数列 $\{h_j(s)\}$ 和分布函数 $F(x)$，有

$$\int_a^b \sum_{j=1}^{\infty} h_j(s)dF(s) = \sum_{j=1}^{\infty} \int_a^b h_j(s)dF(s).$$

定理 4.11: 有界收敛定理

设随机变量序列 $\{X_n\}$ 有界，$|X_n| \le M$ a.s.，$n \ge 1$. 如果 $X_n \to X$ a.s.，则

$$\lim_{n\to\infty} E[X_n] = E[X].$$

又设函数列 $\{h_n(t)\}$ 有界：$|h_n(t)| \le M$. 如果对 $t \in (a,b)$，$h_n(t) \to h(t)$，则在有限区间 $[a,b]$ 上，由

$$\lim_{n\to\infty} \int_a^b h_n(s)ds = \int_a^b h(s)ds.$$

定理 4.12: 勒贝格控制收敛定理

设 $X_1, X_2, \cdots$ 为一列随机变量，且 $X_n \xrightarrow{a.s.} X$. 若存在随机变量 Y，使得 Y 的期望存在且对任意 $n \geq 1$，$|X_n| \leq Y$，则有

$$\lim_{n\to\infty} E[X_n] = E[X].$$

命题 4.13

设随机变量 $X \geq 0$ a.s., $\mu = E[X]$，定义随机变量

$$\tilde{X}_m = \begin{cases} X, & \text{当 } X \leq m \text{ 时}, \\ m, & \text{当 } X > m \text{ 时}, \end{cases}$$

则当 $m \to \infty$ 时，$E[\tilde{X}_m] \to \mu$.

证明.

当 $m \to \infty$ 时，X_m 单调上升趋于 X，于是 $E[\tilde{X}_m] \to \mu$. □

定义 5.1

对于随机变量 $X_1, X_2, \cdots, X_n$，其 **次序统计量**为随机变量 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$，其中

$$
\begin{aligned}
&X_{(1)} = \min(X_1, \cdots, X_n),\\
&X_{(2)} \text{ 是 } X_1, \cdots, X_n \text{ 中第二小的},\\
&\vdots\\
&X_{(n-1)} \text{ 是 } X_1, \cdots, X_n \text{ 中第二大的},\\
&X_{(n)} = \max(X_1, \cdots, X_n).
\end{aligned}
$$

注意到由定义 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. 我们称 $X_{(j)}$ 为 j 阶统计量. 如果 n 是奇数，则称 $X_{(n+1)/2}$ 是 $X_1, \cdots, X_n$ 的样本中位数.

定理 5.2

设 X 有密度函数 $h(x)$，$X_1, X_2, \cdots, X_n$ 是来自于总体 X 的样本，则次序统计量 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 有联合密度

$$g(\boldsymbol{x}) = \begin{cases} n! \prod_{j=1}^{n} h(x_j), & \text{当 } x_1 < x_2 < \cdots < x_n \text{ 时}, \\ 0, & \text{其他}. \end{cases}$$

特别地，当 X 在 $(0, a)$ 上服从均匀分布时，次序统计量 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 有联合密度

$$g(\boldsymbol{x}) = \begin{cases} \dfrac{n!}{a^n}, & \text{当 } 0 < x_1 < x_2 < \cdots < x_n < a \text{ 时}, \\ 0, & \text{其他}. \end{cases}$$