2.4 随机变量的数字特征

2.4.1 期望

很多特征数都涉及到求随机变量或者随机变量函数的期望，期望成为概率论的重要概念。

定义 2.13 (期望) 期望是随机变量取值的加权平均，可表示为: $E(X) = EX$

$$EX = \begin{cases} \sum_i x_i f_X(x_i) \\ \int_{-\infty}^{+\infty} x f_X(x) dx \end{cases}$$

性质 1 随机变量函数的期望: 设连续型随机变量 X, 设随机变量 X 的函数 $g(X)$
(i) 如果 X 是离散型随机变量，则有 $E[g(x)] = \sum_i g(x_i) f_X(x_i)$
(ii) 如果 X 是连续型随机变量，则有 $E[g(x)] = \int g(x) f_X(x) dx$

例 2.13 已知 $X \sim Exp(\lambda)$, 求 $E(X^2)$
$E[Y] = \int_0^{+\infty} y f_Y(y) dy = \int_0^{+\infty} y f_X(x) dx = \int_0^{+\infty} x^2 f_X(x) dx$

性质 2 线性：$E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$

性质 3 设 c 是常数，则有 $E(c) = c$

性质 4 数学期望代表随机变量取值的平均水平，有什么道理？数学期望的统计意义是最小均方误差估计。设随机变量 X,$\forall a \in \mathbb{R}, E[(X-a)^2]$，$a$ 作为 X 的代表的均方误差。

定理： $\min_a E[(X-a)^2] = E[(X-EX)^2]$, 最小均方误差是方差

$$EX = \underset{a}{\operatorname{argmin}}\, E[(X-a)^2]$$

证明 $f(a) = E[(X-a)^2] = E[(X-EX) + (EX-a)]^2 = E[(X-EX)^2] + 0 + (EX-a)^2$

$$\min_a f(a) = E[(X-EX)^2$$

$$\underset{a}{\operatorname{argmin}}\, f(a) = EX$$

2.4.2 方差

方差刻画了随机变量的分散或者波动的程度。

定义 2.14 $Var(X) = E(X-EX)^2 = \sigma_X^2, \sigma_X = \sqrt{Var(X)}$ 称为标准差

性质 1 $Var(X) = E(X^2) - [E(X)]^2$

性质 2 $Var(c) = 0$

性质 3 $Var(aX+b) = a^2 Var(X)$

性质 4 Markov 不等式：设 X 是非负随机变量，则对于 $\forall c > 0, P(X \geq c) \leq \frac{1}{c}E(X)$

证明： $EX = \int_0^{+\infty} xf(x)dx \geq \int_c^{+\infty} xf(x)dx \geq c\int_c^{+\infty} f(x)dx = cP(X \geq c)$

性质 5 Chebyshev 不等式：设 X 均值和方差都存在，则对于 $\forall \varepsilon > 0, P(|X - EX| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}Var(X)$

性质 6 $Var(X) = 0 \Leftrightarrow X$ 几乎处处为某个常数 a, 即 $\exists a, P(x = a) = 1$

证明 欲证 $P(X = EX) = 1$ 等价于证明 $P(X \neq EX) = 0$
$P(X \neq EX) = P(\bigcup_{n=1}^{+\infty}\{|X-EX| \geq \frac{1}{n}\}) \leq \sum_{n=1}^{+\infty} P(|X-EX| \geq \frac{1}{n}) \leq \sum_{n=1}^{+\infty} \frac{1}{(\frac{1}{n})^2}Var(X) = 0$

X 的 k 阶原点矩：$E(X^k)$

X 的 k 阶中心矩：$E[(X - EX)^k]$

期望是一阶原点矩，方差是二阶中心矩。

第三章　多元随机变量

很多随机现象往往涉及多个变量，把这些变量作为一个整体来看。

定义 3.1 从样本空间 Ω 到 $\mathbb{R}^n$ 的一个函数 $X(\omega)$ 称为 n 维随机变量，可表示为
$X(\omega):\Omega\to\mathbb{R}^n$

随机事件的取值表示事件。已知 n 维随机变量 $X\in\mathbb{R}^n$, 数集 $B\subset\mathbb{R}^n$
X 取值于一个数集的随机事件可以表示为:$\{X\in B\}=\{\omega|X(\omega)\in B\}$

例 3.1 随机抽取一个清华同学，

$$X_1(\omega)=\begin{cases}0 & \omega=boy\\ 1 & \omega=girl\end{cases}$$

$X_2(\omega)$: 体重;
$X_3(\omega)$：身高;
$X_4(\omega)$：肺活量
$X_1(\omega),X_2(\omega),X_3(\omega),X_4(\omega)$ 是一个 ω 的不同侧面描述，是一个整体，故这个整体称为 4 元变量。

例 3.2 某邮局有两个柜台，假设当 c 走进邮局时，发现两个职员正在接待 a 和 b，设这时开始 a 和 b 的服务时长为 X_A 和 X_B，于是被告知一旦处理完 a 或者 b 的事情，就立即接待他，c 的服务时长为 X_C, 那么 c 是最后一个办完事情的概率有多大？

解：为了更好地描述这个事件，引入 3 维随机矢量 $X_A(\omega),X_B(\omega),X_C(\omega)$, 那么“$c$ 是最后一个办完事情的概率”可以表示为：$P(min(X_A,X_B)+X_C>max(X_A,X_B))$

定义 3.2 联合累积分布函数 $JCDF$：$\forall x_1,x_2\in R$, 称 $F_{X_1,X_2(x_1,x_2)}=P(X_1\le x_1,X_2\le x_2)$ 为二元变量 (X_1,X_2) 的联合 (累积) 分布函数。

考虑 n 元变量落在以 $(x_1,x_2\cdots x_n)$ 为端点的基本 $Borel$ 集的概率，$(x_1,x_2\cdots x_n)$ 变动起来，得到一个 n 元函数，称为联合 (累积) 分布函数，$JCDF$
$F_{X_1\cdots X_n(x_1\cdots x_n)}=P(X_1\le x_1,X_2\le x_2),\forall x_1\cdots x_n\in\mathbb{R}$

矢量观点　$F_X(x)=P(X\le x),\forall x\in\mathbb{R}^n,X=(X_1,X_2\cdots X_n)^T$

JCDF 的性质　二元随机变量 (X,Y) 的分布函数 $F(X,Y)$ 满足:

性质 1　单调性：$F(X,Y)$ 分别对于 X,Y 单调不减

性质 2　有界性：$\forall x,y,0\leq F(X,Y)\leq 1;F(-\infty,y)=lim_{x\to-\infty}F(X,Y)=0;$
$F(x,-\infty)=lim_{y\to-\infty}F(X,Y)=0;F(+\infty,+\infty)=lim_{x,y\to+\infty}F(X,Y)=1$

性质 3　右连续性：对于每个变量都是右连续的 $F(x+0,y)=F(x,y)$ $F(x,y+0)=F(x,y)$

性质 4　非负性：$P(a\leq X\leq b,c<Y\leq d)>0$

证明　记 $A=\{X\leq a\},B=\{X\leq b\},C=\{Y\leq c\},D=\{Y\leq d\}$ $\bar{A}B\bar{C}D=(BD-AD)-(BC-AC)$
$P(\bar{A}B\bar{C}D)=P(BD-AD)-P(BC-AC)=P(BD)-P(AD)-[P(BC)-P(AC)]=F(b,d)-F(a,d)-F(b,c)+F(a,c)\geq 0$

随机变量取值于数集，一定可以转化为随机变量取值于基本 $Borel$ 集。

定义 3.3 离散多元随机变量的分布列：X 为 n 元离散变量,$X=(X_1\cdots X_n)^T$ 仅取有限个值，每个值都是一个 n 维矢量
联合分布列 $JPMF$: 已知随机变量 $(X,Y)^T$,X 的可能值是 $x_i,i=1,2\cdots$,Y 的可能值是 $y_j,j=1,2\cdots$ 称 $P(X=x_i,Y=y_j)=f(x_i,y_j)$ 为 $(X,Y)^T$ 的联合分布列

性质 1　非负性:$f_X(x_i)\geq 0$

性质 2　正则性:$\sum_i f_X(x_i)=1$

定义 3.4 二元随机变量联合概率密度 $JPDF$：对于二元连续变量, 如果存在二元非负函数 $f_{X,Y}(x,y)$, 使得 $F_{X,Y}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{X,Y}(u,v)dudv$, 则称 $(X,Y)^T$ 为二元连续变量，称 $f_{X,Y}(,)$ 为 $(X,Y)^T$ 的联合密度函数。

性质 1　非负性，$f_(x,y)\geq 0$

性质 2　正则性，$\int_{-\infty}^{x}\int_{-\infty}^{y}f_{X,Y}(u,v)dudv=1$

性质 3　$f_{X,Y}(x,y)$ 是 (x,y) 的连续函数

性质 4　在 $F(x,y)$ 偏导数存在的点上，$f(x,y)=\frac{\partial^2}{\partial x\partial y}F_(x,y)$

性质 5　连续随机矢量在单点上的取值概率等于 0,$P(X=u,Y=v)=0,\forall u,v\in R$

性质 6　微元概率 $P(u<X\leq u+du,v<Y\leq v+dv)=\int_{x,y}(u,v)dxdy$