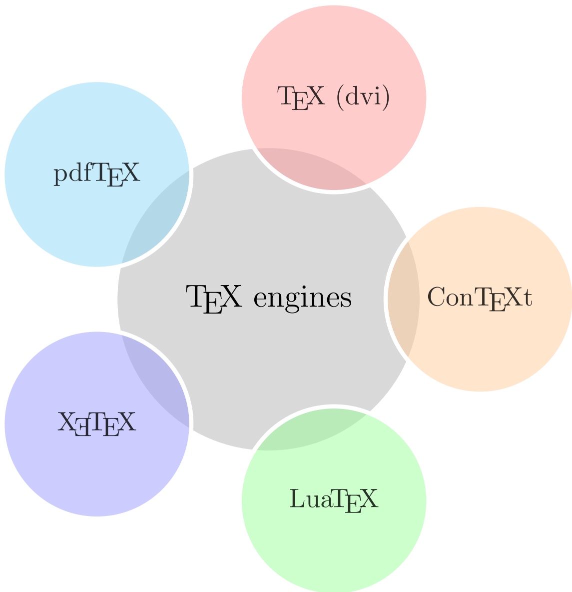
大二上学期复习笔记



adydio
dyk2021@mail.ustc.edu.cn
2022 年 12 月
中国科学技术大学

1 概率论

主要侧重于数字特征与特征函数、极限定理的内容。

1.1 协方差

定义 1 (协方差). $\mathrm{Cov}(X,Y) := \mathrm{E}[(X-\mathrm{E}X)(Y-\mathrm{E}Y)]$.

定理 1.1. 协方差满足如下等式:
(a)$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$;
(b)$\mathrm{Var}(X+Y)=\mathrm{Var}X+2\mathrm{Cov}(X,Y)+\mathrm{Var}Y$;
(c)$\mathrm{Cov}(X,Y)=\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]$;
(d)$\mathrm{Cov}(X,Y)=\mathrm{E}[X(Y-\mathrm{E}Y)]$;
(e)$\mathrm{Cov}(aX,Y)=a\mathrm{Cov}(X,Y)$;
(f) 对 $\forall a_1,a_2,b_1,b_2\in\mathbb{R}$, 有:

$$\mathrm{Cov}(a_1X_1+a_2X_2,b_1Y_1+b_2Y_2)=\sum_{i,j=1,2}a_ib_j\mathrm{Cov}(X_i,Y_j)$$

一个例子是, 对于随机变量 X,Y 满足二维正态分布 $\mathcal{N}(a,b;\sigma_1,\sigma_2;r)$,

$$\mathrm{Cov}(X,Y)=r\sigma_1\sigma_2.$$

1.2 条件期望

定义 2 (条件期望). 当 X 和 Y 的联合分布为离散分布时, 在给定 $Y=y$ 之下, X 的条件分布列定义为:

$$p_{X|Y}(x|y)=P(X=x|Y=y)=\frac{p(x,y)}{p_Y(y)}.$$

此时, X 在给定 $Y=y$ 之下的条件期望为:

$$\mathrm{E}[X|Y=y]=\sum_x xp_{X|Y}(x|y).$$

类似地，当 X 和 Y 的联合分布为连续分布时，在给定 $Y=y$ 之下，$f_Y(y)>0$ 时，X 的条件密度函数定义为：

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}.$$

自然地，在给定 $Y=y$ 条件下，X 的条件期望为：

$$\mathrm{E}[X|Y=y]=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)dx.$$

条件期望带来了一个比较直观的公式：

定理 1.2. $\mathrm{E}[X]=\mathrm{E}[\mathrm{E}[X|Y]]$.

$\mathrm{E}[X|Y]$ 是与 Y 有关的随机变量，可以取期望。这个公式提供了求数学期望“两步走”的方法。

定义 3 (条件方差). 定义 $Y=y$ 条件之下的条件方差：

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

定理 1.3 (条件方差公式).

$$\mathrm{Var}(X)=\mathrm{E}[\mathrm{Var}(X|Y)]+\mathrm{Var}(\mathrm{E}[X|Y]).$$

证明. 类似于方差，我们有：

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

注意到，这时候 $\mathrm{Var}(X|Y)$ 是一个与 Y 相关联的随机变量，它也是存在期望的。对上式两端取极限，得到：

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

同时也可以观察 $\mathrm{Var}(\mathrm{E}[X|Y])$ 的展开：

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

相加上面两式即证。 □

利用期望和条件期望等概念可以引出很多附带的结论。作为例子给出：

例 1. $\mathrm{E}[(X-a)^2]$ 在 $a=\mathrm{E}[X]$ 时取到最小。如果利用随机变量 X 的值来对另一个随机变量 Y 进行预测，那么 Y 的最优预测值为 $g(X)=\mathrm{E}[Y|X]$，即：

$$\mathrm{E}[(Y-g(X))^2]\geq\mathrm{E}[(Y-\mathrm{E}[Y|X])^2].$$

例 2. $r = \mathrm{Cov}(X, Y)$，证明性质：

$$\mathrm{E}[\mathrm{Var}(Y|X)] \leq (1 - r^2)\mathrm{Var}(X).$$

证明. 在证明只含有方差的结论时，可不妨设研究对象的均值为零。这里令 $\mathrm{E}X = \mathrm{E}Y = 0$。

$$r^2\mathrm{Var}(X)\mathrm{Var}(Y) = \mathrm{E}[XY]^2 = \mathrm{E}[X\mathrm{E}[Y|X]]^2 \leq \mathrm{E}[X^2]\mathrm{E}[\mathrm{E}[Y|X]^2]$$

再由条件方差公式，得到：

$$\mathrm{E}[\mathrm{Var}(Y|X)] = \mathrm{Var}(Y) - \mathrm{Var}(\mathrm{E}[Y|X]) \leq (1 - r^2)\mathrm{Var}(Y).$$

□

1.3 几种重要分布

定义 4 (**几何分布**)**.** 设有一独立重复试验序列，每次试验成功概率为 p。记 N 为取得第一次成功所需要的试验次数，则 N 服从几何分布，记为 $N \sim Geo(p)$。N 满足：

$$P(N = x) = p(1 - p)^{x-1}.$$

对于 $N \sim Geo(p)$，其均值和方差为：

$$\mathrm{E}[N] = \frac{1}{p};$$

$$\mathrm{Var}(N) = \frac{1 - p}{p^2}.$$

下面证明之：

证明. 第一次成功，$Y = 1$，否则取 0。

$$\mathrm{E}[N] = \mathrm{E}[\mathrm{E}[N|Y]] = \mathrm{E}[p \cdot 1 + (1 - p)(\mathrm{E}(N) + 1)] \Rightarrow \mathrm{E}[N] = \frac{1}{p}.$$

$$\mathrm{E}[N^2] = \mathrm{E}[N^2|Y = 1]P(Y = 1) + \mathrm{E}[N^2|Y = 0]P(Y = 0) \Rightarrow \mathrm{E}[N^2] = 1 + (1 - p)\mathrm{E}[N^2 + 2N]$$

$$\Rightarrow \mathrm{E}[N^2] = \frac{2 - p}{p^2} \Rightarrow \mathrm{Var}(N) = \frac{1 - p}{p^2}.$$

□

定义 5 (B **分布**)**.** B 分布是定义在 $(0, 1)$ 上的连续概率分布，其概率密度为：

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}.$$

B 分布的期望和方差为：

$$\mathrm{E}X = \frac{\alpha}{\alpha + \beta};$$

$$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

1.4 特征函数

定义 6 (特征函数). $F(x)$ 为一维分布函数，将 $f(t)=\int_{-\infty}^{\infty}e^{itx}dF(x)$ 称为 $F(x)$ 的特征函数。如果 $F(x)$ 是随机变量 X 的分布函数，则 $f(t)$ 也称为随机变量 X 的特征函数，有 $f(t)=\mathrm{E}e^{itx}$。

一些常见的分布函数需要熟练推导，作为例子给出：

例 3 (退化于 a 的随机变量). $f(t)=e^{ita}$。

例 4 (参数为 p 的 Bernoulli 分布). $f(t)=pe^{it}+q$。

例 5 (两点分布). 对于满足两点分布(取 a 的概率为 p，取 b 的概率为 $q=1-p$)的随机变量，特征函数为 $f(t)=pe^{ita}+qe^{itb}$。特别地，$a=-b=1;p=\frac{1}{2}$ 时，$f(t)=\frac{e^{it}+e^{-it}}{2}=\mathrm{cost}$。

例 6 (参数为 λ 的 Poisson 分布). 对于 Poisson 分布，$P(X=n)=e^{-\lambda}\frac{\lambda^n}{n!}$，$f(x)=e^{-\lambda}\sum_{n=1}^{\infty}\frac{(e^{it}\lambda)^n}{n!}=e^{\lambda(e^{it}-1)}$。

例 7 (参数为 p 的几何分布). $f(t)=\sum_{n=1}^{\infty}e^{itn}pq^{n-1}=\frac{pe^it}{1-qe^{it}}$。

例 8 (参数为 λ 的指数分布). $f(t)=\frac{\lambda}{\lambda-it}$。注意，在这里直接积分会出问题，应该把实部和虚部分来来积。

分布函数的特征函数具有一些性质：

定理 1.4. (a)$|f(t)|\leq f(0)=1,\forall t\in\mathbb{R}$;
(b)$f(-t)=\overline{f(t)}$;
(c)$f(t)$ 在 $\mathbb{R}$ 上一致连续;
(d)$f_{a+bX}(t)=e^{ita}f_X(bt)$.

1.5 多维正态分布

定义 7 (二维正态分布). 随机变量分布叫做二维正态分布，如果它的概率密度函数满足：

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp[-\frac{1}{2(1-\rho^2)}(\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2})].$$

二维正态分布的边际分布也是正态分布。即：

$$f_1(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp(-\frac{(x-\mu_1)^2}{2\sigma_1^2}).$$

此外，随机变量 X,Y 相互独立 $\Leftrightarrow\rho=0$。

例 9. 证明对于满足二维正态分布的随机变量 X, Y，它们的期望均为零，有：

$$\mathrm{E}[X|Y] = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} Y.$$

证明. □

1.6 极限定理

定理 1.5 (**马尔可夫不等式**)**.** 设 X 为取非负值的随机变量，则对于任何常数 $a > 0$，有：

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

证明. 取事件 $X \geq a$ 的示性变量 I, 则有 $I \leq \frac{X}{a}$, 两边取期望, 得到 $P(X \geq a) = \mathrm{E}[I] \leq \frac{\mathrm{E}[X]}{a}$ 得证。 □

定理 1.6 (**切比雪夫不等式**)**.** 设 X 为随机变量, 均值 μ 和方差 σ^2 有限, 则对任何 $k > 0$, 有：

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

证明.

$$P(|X - \mu| \geq k) = P((X - \mu)^2 \geq k^2) \leq \frac{\mathrm{E}[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

上述证明用到了马尔可夫不等式。 □

定理 1.7 (**切比雪夫不等式的推广**)**.** 推而广之，我们有 $g(x)$ 为定义在 $[0, +\infty)$ 上的非降的非负值函数，对随机变量 X，如果有 $\mathrm{E}g(|X|) < \infty$ 则对于任意使 $g(a) > 0$ 的 $a > 0$，都有：

$$P(|X| > a) \leq \frac{\mathrm{E}g(|X|)}{g(a)}.$$

定理 1.8 (**弱大数定律**)**.** 设 $X_1, X_2 ... iid$，$\mathrm{E}[X_1] = \mu < \infty$，则对任意 $\varepsilon > 0$，

$$S_n = \frac{X_1 + ... + X_n}{n} \xrightarrow{p} \mu.$$

注意，这里没有要求 X_1 的方差存在。