数理统计第一次中期测验

摘　　　　要

该读书报告聚焦于概率统计和计算数学中常用的概率分布，并整理了它们的密度函数、数学期望与方差、特征函数和应用场合，是对概率论的复习和数理统计课本以外的学习.

笔者分别在第一部分整理了常用的离散分布，包括 Bernoulli 分布，二项分布，Possion 分布，超几何分布，几何分布，Pascal 分布，负二项分布等；在第二部分整理了常见的连续分布，包括均匀分布，正态分布，Laplace 分布，Rayleigh 分布，指数分布，Γ 分布 (Erlang 分布,Wishart 分布)，Cauchy 分布，β 分布 (Dirichlet 分布)，Pareto 分布，广义 Pareto 分布，Levy 分布，对数正态分布，广义极值分布 (Gumbel 分布，Frechet 分布，Weibull 分布) 等；在第三部分专门整理了统计学中的三大分布，即 χ^2 分布，t 分布，F 分布.

离散分布

Bernoulli 分布 $X \sim B(1, p)$

密度函数 $P(x=k) = \begin{cases} 1-p, & p=0 \\ p, & k=1 \end{cases}$, $(0<p<1)$;

数学期望 $EX = p$; 方差 $DX = p(1-p)$; 特征函数 $f(t) = pe^{it} + (1-p)$.

应用: 抛一次硬币模型; 实现二元分类.

二项分布 $X \sim B(n, p)$

密度函数 $P(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$, $(k=0, \cdots, n,\ 0<p<1)$;

数学期望 $EX = np$; 方差 $DX = np(1-p)$; 特征函数 $f(t) = \left(pe^{it} + (1-p)\right)^n$.

应用: 多次抛硬币模型.

Possion 分布 $X \sim P(\lambda)$

密度函数 $P(x=k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$, $(k \in \mathbb{N},\ \lambda > 0)$;

数学期望 $EX = \lambda$; 方差 $DX = \lambda$; 特征函数 $f(t) = \exp\left(\lambda\left(e^{it} - 1\right)\right)$.

应用: 某段连续的时间内某件事情发生的次数; Possion 回归模型.

超几何分布 $X \sim H(N, n, M)$

密度函数 $P(x=k) = \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, $(M, n \le N,\ n, M, N \in \mathbb{N}_+,\ 0 \le k \le \min\{M, n\})$;

数学期望 $EX = \dfrac{nM}{N}$; 方差 $DX = \dfrac{nM}{N}\left(1 - \dfrac{M}{N}\right)\dfrac{N-n}{N-1}$;

特征函数 $f(t) = \displaystyle\sum_{k=0}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} e^{itk}$.

应用: 随即取出次品数; N 很大时近似为二项分布.

几何分布 $X \sim G(p)$

密度函数 $P(x=k) = (1-p)^{k-1} p, \, (k \in \mathbb{N}_+, \, 0 < p < 1);$

数学期望 $EX = \dfrac{1}{p}$; 方差 $DX = \dfrac{1-p}{p^2}$; 特征函数 $f(t) = \dfrac{pe^{it}}{1-(1-p)e^{it}}$.

应用: 成功概率为 p 的 Bernoulli 实验中, 首次成功的试验次数.

Pascal 分布

密度函数 $P(x=k) = \dbinom{k-1}{r-1} p^r (1-p)^{k-r}, \, (k = r, r+1, \cdots, \, 0 < p < 1);$

数学期望 $EX = \dfrac{r}{p}$; 方差 $DX = \dfrac{r(1-p)}{p^2}$; 特征函数 $f(t) = \left(\dfrac{pe^{it}}{1-(1-p)e^{it}} \right)^r$.

应用: 成功概率为 p 的 Bernoulli 实验中, 成功第 r 次时的试验次数.

负二项分布 $X \sim NB(r,p)$

密度函数 $P(x=k) = p\dbinom{-r}{k} p^r (p-1)^k, \, (k \in \mathbb{N}, \, 0 < p < 1, \, r > 0);$

数学期望 $EX = \dfrac{r(1-p)}{p}$; 方差 $DX = \dfrac{r(1-p)}{p^2}$; 特征函数 $f(t) = \left(\dfrac{p}{1-(1-p)e^{it}} \right)^r$.

应用: 在生物重复中计数统计; 在 r 很大时是 Possion 分布的近似.

连续分布

均匀分布 $X \sim U[a,b]$

密度函数 $P(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{others} \end{cases}$, $(a < b \in \mathbb{R})$;

数学期望 $EX = \dfrac{a+b}{2}$; 方差 $DX = \dfrac{(b-a)^2}{12}$; 特征函数 $f(t) = \dfrac{e^{itb} - e^{ita}}{it(b-a)}$.

应用: 等概率取点.

正态分布 $X \sim N(\mu, \sigma^2)$

密度函数 $P(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$, $(x \in \mathbb{R},\ \mu, \sigma > 0)$;

数学期望 $EX = \mu$; 方差 $DX = \sigma^2$; 特征函数 $f(t) = \exp\left(i\mu t - \dfrac{1}{2}\sigma^2 t^2\right)$.

应用: 市场营销中预测次年购买趋势; 金融业刻画高频交易的波动特征; 疾病的早期警示; 保险的盈亏概率; 根据消费者特征进行风险管控; 预测股票涨跌等.

Laplace 分布

密度函数 $P(x) = \dfrac{1}{2\lambda} \cdot \exp\left(-\dfrac{|x-\mu|}{\lambda}\right)$, $(\mu \in \mathbb{R},\ \lambda > 0)$;

数学期望 $EX = \mu$; 方差 $DX = 2b^2$; 特征函数 $f(t) = \dfrac{e^{i\mu t}}{1 + \lambda^2 t^2}$.

应用: 刻画证券金融交易; 用于测绘数据的处理以及在语音和图像数据等领域.

Rayleigh 分布

密度函数 $P(x) = \dfrac{x}{\sigma^2} \cdot \exp\left(-\dfrac{x^2}{2\sigma^2}\right)$, $(x > 0,\ \sigma > 0)$;

数学期望 $EX = \sigma\sqrt{\dfrac{\pi}{2}}$; 方差 $DX = \dfrac{4-\pi}{2}\sigma^2$; 特征函数 $f(t) = \dfrac{e^{i\mu t}}{1 + \lambda^2 t^2}$.

应用: 最早用于预测逃逸到客户现场的缺陷情况; 是软件工程学中双向质量控制策略的重要工具, 可以对软件开发全生命周期进行预测, 也可以仅对测试阶段的缺陷分布进行预测.

指数分布 $X \sim \exp(\lambda)$

密度函数 $P(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}, (\lambda > 0);$

数学期望 $EX = \dfrac{1}{\lambda}$; 方差 $DX = \dfrac{1}{\lambda^2}$; 特征函数 $f(t) = \left(1 - \dfrac{it}{\lambda}\right)^{-1}$.

应用: 作用一般体现在等到一个随机事件发生需要经历的时间; 常用来描述寿命类随机变量的分布, 如家电使用寿命, 动植物寿命, 通话时间等.

Γ 分布 $X \sim \Gamma(r, \lambda)$

密度函数 $P(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}, (r > 0,\ \lambda > 0);$

数学期望 $EX = \dfrac{r}{\lambda}$; 方差 $DX = \dfrac{r}{\lambda^2}$; 特征函数 $f(t) = \left(1 - \dfrac{it}{\lambda}\right)^{-r}$.

应用: $r \in \mathbb{N}_+$ 时称作 Erlang 分布; 一般多元情况称作 Wishart 分布. 作用一般体现在等到 n 个随机事件都发生需要经历的时间, 在 Possion 过程中应用, 表现为多个独立且相同分布的指数分布变量的和的分布.

Cauchy 分布

密度函数 $P(x) = \dfrac{1}{\pi} \dfrac{\lambda}{\lambda^2 + (x - \mu)^2}, (\lambda > 0,\ \mu \in \mathbb{R});$

数学期望不存在 ($EX = \infty$); 方差不存在 ($DX = \infty$); 特征函数 $f(t) = \exp(i\mu t - \lambda|t|)$.

应用: 用于机械和电气理论的测量和校准问题; 用于金融模型中表示预测模型的收益偏差; 是量子力学中不稳定态能量的分布, 也用于对从点源发出的固定直线的粒子的碰撞点进行建模.

β 分布

密度函数 $P(x) = \begin{cases} \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1}, & 0 < x < 1 \\ 0 & \text{others} \end{cases}, (p, q > 0);$

数学期望 $EX = \dfrac{p}{p+q}$; 方差 $DX = \dfrac{pq}{(p+q)^2 (p+q+1)}$;

特征函数 $f(t) = \dfrac{\Gamma(p+q)}{\Gamma(p)} \displaystyle\sum_{j=0}^{\infty} \dfrac{\Gamma(p+j)(it)^j}{\Gamma(p+q+j)\Gamma(j+1)}$.

应用: 描述随机的百分比; 一般多元形式称作 Dirichlet 分布.

Pareto 分布

密度函数 $P\left(x\right)=\begin{cases}\dfrac{\alpha A^{\alpha}}{A^{\alpha+1}}, & x\geq A\\ 0, & x<A\end{cases}, \left(\alpha>0,\ A>0\right);$

数学期望 $EX=\dfrac{\alpha A}{\alpha-1}, \left(\alpha>1\right)$; 方差 $DX=\dfrac{\alpha A^{2}}{\left(\alpha-1\right)^{2}\left(\alpha-2\right)}, \left(\alpha>2\right);$

特征函数 $f\left(t\right)=\alpha\left(-iAt\right)^{\alpha}\Gamma\left(-\alpha,-iA\right)$.

应用: 描述社会财富的分布 (二八法则), 也被用于地理学和精算学.

广义 Pareto 分布 $X\sim GPD\left(\mu,\sigma,\xi\right)$

密度函数 $P\left(x\right)=\dfrac{1}{\sigma}\left(1+\dfrac{\xi\left(x-\mu\right)}{\sigma}\right)^{-1-\xi^{-1}}, \left(\mu,\xi\in\mathbb{R},\ \sigma>0\right);$

数学期望 $EX=\mu+\dfrac{\sigma}{1-\xi}, \left(\xi<1\right)$; 方差 $DX=\dfrac{\sigma^{2}}{\left(1-\xi\right)^{2}\left(1-2\xi\right)}, \left(\xi<\dfrac{1}{2}\right);$

特征函数 $f\left(t\right)=e\left(i\mu t\right)\sum\limits_{j=0}^{\infty}\dfrac{\left(i\sigma t\right)^{j}}{\prod\limits_{k=0}^{j}\left(1-k\xi\right)}, \left(k\xi<1\right)$.

应用: 描述其他分布的尾部分布.

Levy 分布

密度函数 $P\left(x\right)=\sqrt{\dfrac{c}{2\pi}}\dfrac{\exp\left(-\frac{c}{2\left(x-\mu\right)}\right)}{\left(x-\mu\right)^{\frac{3}{2}}}, \left(x\geq\mu,\ c>0\right);$

数学期望不存在 $\left(EX=\infty\right)$; 方差不存在 $\left(DX=\infty\right)$; 特征函数 $f\left(t\right)=\exp\left(i\mu t-\sqrt{-2ict}\right)$.

应用: 在风险管理和套利策略设计等方面应用广泛.

对数正态分布

密度函数 $P\left(x\right)=\begin{cases}\dfrac{1}{\sigma x\sqrt{2\pi}}\exp\left(-\dfrac{\left(\ln x-a\right)^{2}}{2\sigma^{2}}\right), & x>0\\ 0, & x\leq 0\end{cases}, \left(a\in\mathbb{R},\ \sigma>0\right);$

数学期望 $EX=\exp\left(\alpha+\dfrac{\sigma^{2}}{2}\right)$; 方差 $DX=\left(e^{\sigma^{2}}-1\right)\exp\left(2\alpha+\sigma^{2}\right)$.

应用; 短期来看与正态分布非常接近, 但长期来看有更大向上波动的可能和更小向下波动的可能, 常用于电子元件的可靠性分析和某些种类的机械零件的疲劳寿命等.

广义极值分布 $X \sim GEV\left(\mu,\sigma,\xi\right)$

密度函数 $P\left(x\right)=\dfrac{1}{\sigma}t\left(x\right)^{\xi+1}e^{-t(x)}$, $t\left(x\right)=\begin{cases}\left(1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}, & \xi\neq 0\\ e^{\frac{\mu-x}{\sigma}}, & \xi=0\end{cases}$, $(\mu,\xi\in\mathbb{R},\,\sigma>0)$;

数学期望 $EX=\begin{cases}\dfrac{\mu+\sigma\left(\Gamma\left(1-\xi\right)-1\right)}{\xi}, & \xi\neq 0,\xi<1\\ \mu+\sigma\displaystyle\int_1^{\infty}\left(-\frac{1}{x}+\frac{1}{\lfloor x\rfloor}\right)\mathrm{d}x, & \xi=0\\ \infty, & \xi\geq 1\end{cases}$;

方差 $DX=\begin{cases}\dfrac{\sigma^2\left(\Gamma\left(1-2\xi\right)-\Gamma\left(1-\xi\right)^2\right)}{\xi^2}, & \xi\neq 0,\xi<\frac{1}{2}\\ \sigma^2\dfrac{\pi^2}{6}, & \xi=0\\ \infty, & \xi\geq\frac{1}{2}\end{cases}$.

应用: 在 $\xi=0$ 时称作 Gumbel 分布, 一般用于预测严重的自然灾害, 如地震和洪水; $\xi>0$ 时称作 Frechet 分布; $\xi<0$ 时称作 Weibull 分布.

三大统计分布

χ^2 分布 $X \sim \chi^2(n)$

$$\text{密度函数 } \chi^2(x,n) = \begin{cases} \dfrac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}, (n \in \mathbb{N}_+);$$

数学期望 $EX = n$; 方差 $DX = 2n$; 特征函数 $f(t) = (1-2it)^{-\frac{n}{2}}$.

应用: 单正态总体方差的假设检验, 总体分布未知情况下的分布拟合检验等.

t 分布 $X \sim t(n)$

$$\text{密度函数 } t(x,n) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, (n \in \mathbb{N}_+);$$

数学期望 $EX = 0, (n > 1)$; 方差 $DX = \dfrac{n}{n-2}, (n > 2)$.

应用: 用来检测不同的组之间的均值是否相等, 如正态总体均值假设检验.

F 分布 $X \sim F(p,q)$

$$\text{密度函数 } P(x,p,q) = \begin{cases} \dfrac{\Gamma\left(\dfrac{p+q}{2}\right)}{\Gamma\left(\dfrac{p}{2}\right)\Gamma\left(\dfrac{q}{2}\right)}p^{\frac{p}{2}}q^{\frac{q}{2}} \cdot \dfrac{x^{\frac{p}{2}-1}}{(q+px)^{\frac{p+q}{2}}}, & x \geq 0 \\ 0, & x < 0 \end{cases}, (p,q \in \mathbb{N}_+);$$

数学期望 $EX = \dfrac{q}{q-2}, (q > 2)$; 方差 $DX = \dfrac{2q^2(p+q-2)}{p(q-2)^2(q-4)}, (q > 4)$.

应用: 用来检测不同组之间的方差是否相等, 如多正态总体方差的假设检验.