统计 笔记

任云玮

目录

1 常见分布

1.1 离散分布

1.1.1 Hypergeometric Distribution

含义 设有 N 个球，其中 M 个为红色，$N-M$ 个为绿色，它们除颜色以外无区别. 从中选取 K 个，考虑恰有 x 个是红球的概率分布.

公式

$$
\begin{aligned}
&P(X=x|N,M,K)=\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}.\\
&\operatorname{E}X=\frac{KM}{N}.\\
&\operatorname{Var}X=\frac{KM}{N}\frac{(N-M)(N-K)}{N(N-1)}.\\
&x=0,1,\ldots,K.
\end{aligned}
$$

1.1.2 Binomial Distribution

含义 考虑 n 次相同的成功概率为 p 的 Bernoulli 试验，考虑其中恰有 y 次成功的概率分布.

公式

$$
\begin{aligned}
&P(Y=y|n,p)=\binom{n}{y}p^y(1-p)^{n-y}.\\
&\operatorname{E}X=np\\
&\operatorname{Var}X=np(1-p).\\
&M_X(t)=[pe^t+(1-p)]^n.
\end{aligned}
$$

1.1.3 Poisson Distribution

含义 Poisson 分布可以用来描述在一段时间内，某事件发生的次数的概率分布，例如类似于等车的行为. 设 $t\ge 0$ 而 N_t 是关于 t 的整数值随机变量，设它满足如下性质：

1. 初始无到达：$N_0=0$；
2. 不相交时间区间内的达到次数相互独立：设 $s<t$，则 N_s 和 N_t-N_s 独立；
3. 到达次数近于区间长度有关：N_s 和 $N_{t+s}-N_t$ 是同分布的；
4. 当时间区间充分小时，到达可能性正比于区间长度：$\lim_{t\to 0}\dfrac{P(N_t=1)}{t}=\lambda$；

5. 不会出现同时到达的情况：$\lim_{t\to 0}\dfrac{P(N_t > 1)}{t} = 0$；

经常的，我们会进行归一化，同时只考虑 $t = 1$ 的情况，即"仅等一单位时间情况下，等到车的数量".

公式

$$\begin{aligned}
&P(N_t = n|\lambda) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.\\
&\mathrm{E}\, N_1 = \lambda.\\
&\mathrm{Var}\, N_1 = \lambda.\\
&M_{N_1}(t) = e^{\lambda(e^t-1)}.
\end{aligned}$$

推导 我们仅考虑归一化的结果，即一个单位时间的情况. 我们把该单位时间等分成足够多的 n，则对于每一个小时间段，可以认为互相独立的 Bernoulli 试验，其成功的概率是 λ/n. 则该单位时间的到达次数，即为这 n 次 Bernoulli 试验的成功次数，即为一个二次分布，我们令 $n \to \infty$，即得到 Poisson 分布的 pmf，

$$f(x) = \lim_{n\to\infty}\binom{n}{x}(\lambda/n)^x(1-\lambda/n)^{1-x} = e^{-\lambda}\frac{\lambda^x}{x!}.$$

性质

1 **定理** 设 $X \sim \text{Poisson}(\theta)$，$Y \sim \text{Poisson}(\lambda)$ 且 X 和 Y 独立，则 $X + Y \sim \text{Poisson}(\theta + \lambda)$.

利用 Poisson 分布近似 注意到 Poisson 描述的是连续的时间段中事件发生的情况，在某些离散的情况下，亦可以使用 Poisson 分布来近似. 考虑二项分布的情况，我们可以把 Bernoulli 试验的成功理解成 Poisson 分布中的"到达"，同时把单个的 Bernoulli 试验理解成一个单位时间，则自然的 p 就对应着 λ. 若 p 充分小，则可以认为这些离散的 Bernoulli 试验已经和连续的情况差不多乐，所以此时可以有

$$P(Y = y|n, p) \sim P(N_n = y|p) = P(N_1 = y|np).$$

我们可以通过比较它们的递推式来证明这一结论.

1.1.4 Negative Binomial Distribution

含义 考虑 n 次独立的 Bernoulli(p) 试验，设随机变量 X 表示在第 X 轮发生了等 r 次成功，即描述一个成功了 r 次后才会停止的试验. 或者考虑一种等价的形式，设随机变量 Y 表示在第 r 次成功前的失败的次数，显然有 $Y = X - r$.

公式

$$
\begin{aligned}
&P(X=x|r,p)=\binom{x-1}{r-1}p^r(1-p)^{x-r}\\
&P(Y=y|r,p)=\binom{r+y-1}{y}p^r(1-p)^y=(-1)^y\binom{-r}{y}p^r(1-p)^y.\\
&\mathrm{E}\,Y=r\frac{1-p}{p}.\\
&\mathrm{Var}\,Y=r\frac{1-p^2}{p}=\mu+\frac{1}{2}\mu^2.
\end{aligned}
$$

利用负二项分布近似 Poisson 分布 注意到如果我们令 $p\to1$ 而 $r\to\infty$ 且同时满足 $r(1-p)\to\lambda$，则期望与方差都会趋于 λ，和 Poisson 分布的结果一致. 这暗示了在这一条件下，负二项分布或许可以近似 Poisson 分布，而事实确实如此.

我们可以这样考虑这个事情. 当 r 变大时候，一次试验在其中所占的份就变小了，从而变得更加的连续了，而在这里把一次失败理解成一次到达，则 $p\to1$ 的条件同 Poisson 分布中关于概率与区间长度的要求是一致的，而 $r(1-p)\to\lambda$ 的要求可以从归一化之后的角度看待.

1.1.5 Geometric Distribution

含义 负二项分布中，取 $r=1$ 的特殊情况，即一种成功就停止的试验.

公式

$$
\begin{aligned}
&P(X=x|p)=p(1-p)^{x-1}.\\
&\mathrm{E}\,X=\frac{1}{p}.\\
&\mathrm{Var}\,X=\frac{1-p}{p^2}.
\end{aligned}
$$

性质 之前的失败对于之后的概率没有影响，即有

$$P(X>s|X>t)=P(X>s-t).$$

这也意味着如果某种概率是随着试验次数/时间的增长而有变化时，几何分布是不适用的.

1.2 连续分布

1.2.1 Exponential Distribution

含义 已知在单位时间内事件发生的次数为 λ，W 为第一次事件发生前所经过的时间.

推导 我们下求 W 的 cdf 并求导以得其 pdf. 设 P_λ 为 Poisson 分布，则

$$\begin{aligned} &F(x) = 1 - P(W > x) = 1 - P_\lambda(N_x = 0) = 1 - e^{-\lambda x} \\ \Rightarrow \quad &f(x) = F'(x) = \lambda e^{-\lambda x}. \end{aligned}$$

公式 设 $\theta = 1/\lambda$ 表示所需等时间的期望，则

$$\begin{aligned} &f(x) = \frac{1}{\theta} e^{-x/\theta}, \\ &\mathrm{E}(x) = \theta, \\ &\mathrm{Var}(x) = \theta^2, \\ &x \geq 0, \theta > 0. \end{aligned}$$

1.2.2 Gamma Distribution

含义 设单位时间内事件发生的次数为 λ，W 为第 α 次事件发生的时间. 记 $\theta = 1/\lambda$.

公式

$$\begin{aligned} &f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \\ &E(W) = \alpha\theta \\ &\mathrm{Var}(W) = \alpha\theta^2 \\ &x \geq 0, \alpha, \theta > 0. \end{aligned}$$

1.2.3 正态分布

公式 其参数即为其均值 μ 和方差 σ^2.

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty.$$

若 $X \sim \mathrm{n}(\mu, \sigma^2)$，则称 $Z = (X - \mu)/\sigma \sim n(0, 1)$ 呈 **标准正态分布**.

性质 正态分布的随机变量的 pdf 在 $x = \mu$ 处达到最大值，在 $x = \mu \pm \sigma$ 处变换凹凸性.

利用正态分布近似二项分布 设 $X \sim \mathrm{binomial}(n, p)$ 且 $Y \sim \mathrm{n}(np, np(1-p))$，则

$$P(X \leq x) \approx P(Y \leq x + 1/2), \quad P(X \geq x) \approx P(Y \geq x - 1/2).$$

其中 $\pm 1/2$ 为连续性修正.

2 Cheat Sheet

2 引理 (独立) 设 (X, Y) 为二元随机向量的 pmf 为 $f(x, y)$. 则随机变量 X 和 Y 独立当且仅当存在 $g(x)$ 和 $h(y)$ 使得对任意 $x, y \in \mathbb{R}$ 成立 $f(x, y) = g(x)h(y)$.

3 定理 (变换的 pdf) 设 $(U, V) = g(X, Y)$ 是从 $\mathcal{A} = \{(x, y) : f_{X,Y}(x, y) > 0\}$ 到 $\mathcal{B} = g(\mathcal{A})$ 的双射. 设 $h = g^{-1}$ 而 J 是 h 的 Jacob 行列式且 J 不恒为 0，则

$$f_{U,V}(u, v) = f_{X,Y}(h(u, v))|J|.$$

TODO: p161

4 引理 (独立) 设 X 和 Y 为两个独立随机变量且 g 和 h 分别是仅关于 x 和 y 的函数，则随机变量 $U = g(X)$ 和 $V = h(Y)$ 也独立.

5 引理 (协方差) $\operatorname{Cov}(X, Y) = \mathrm{E}((X - \mu_X)(Y - \mu_Y)) = \mathrm{E}\, XY - \mu_x \mu_Y$.

6 引理 设 $X_1, \ldots, X_n$ 是随机采样且 $g(x)$ 是一个满足 $\mathrm{E}\, g(X_1)$ 和 $\operatorname{Var} g(X_1)$ 存在的函数，则有

$$\mathrm{E}\left(\sum_{i=1}^{n} g(X_i)\right) = n(\mathrm{E}\, g(X_1)),$$
$$\operatorname{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\left(\operatorname{Var} g(X_1)\right).$$

这一引理的 lhs 中的变换的形式（求和），是常见的统计的形式，因此此引理可用于分析统计的均值和方差.

3 注解

p212 Definition 5.2.2/3 **随机变量**是指一个从样本空间 S 到 $\mathbb{R}$ 的映射，而一个随机采样则是 n 个 iid 的随机变量. 而一个**统计**是指一个从一个随机采样到 $\mathbb{R}^n$ 的映射. 而我们所定义的**统计均值**和**统计方差**都是一个统计，即是一个从映射（随机变量）到 $\mathbb{R}^n$ 的映射. 如果我们取定随机采样，则可以认为它是一个从样本空间到 $\mathbb{R}^n$ 的映射，即它也是一个随机变量. 从而我们可以讨论它的均值或者方差，讨论它的分布. 这里要注意，统计均值/方差和均值/方差是完全不同的东西，前者（在取定随机采样后）是一个随机变量，而后者则是一个常数.

p214 Theorem 5.2.6 这里给出了之所以在定义统计方差时，分母用 $n-1$ 而非 n. 只有这样才可以使得 S^2 的期望恰好为 σ^2——各随机变量的各自的方差.

p232 Definition 5.5.1 首先再次注意随机变量是一个从样本空间到 $\mathbb{R}$ 的映射，而 $P(|X_n - X| \geq \varepsilon)$ 表示这样一件事情：如果对于样本空间里的一个元素，如果 X_n 和 X 之间的差不小于 ε，则把这个元素"对应"的概率加上，把所有这些元素都加上的结果即为 $P(|X_n - X| \geq \varepsilon)$. 而此定义表明当 n 充分大的时候，这些概率累加的结果可以充分小.

p234 Definition 5.5.6 对于一个具体的式子，这一定义和 Definition 5.5.1 的一个区别在于，在后者中如果我们展开极限以外的部分，我们会得到一个关于 ε 和 n 的函数，需要这个函数在 $n \to \infty$ 时候趋于零（或者 1）. 而在这一定义中，我们将不得不先处理极限，然后考察不满足要求的样本集合对应的概率. 这一定义相较于 5.5.1 要更强，它表示函数"从概率角度"几乎处处点态收敛，而在 5.5.1 中，它实际上是一个数列的收敛性，可以理解为它是某个"概括"了 X_n 的数量的收敛性。

p235 Definition 5.5.10 即对应的 cdf 列在 F_X 的连续点集合上点态收敛。

p236 Theorem 5.5.13 "converges in distribution to μ" 的直观含义就是结果一定是 μ。

p236 Theorem 5.5.14 这一定理表明在一定条件下，$\sqrt{n}(\bar{X}_n - \mu)/\sigma$ 趋于 $n(0,1)$. 这一定理的作用在于，假设我们从一个分布中进行随机采样，则当样本足够多时，样本均值的分布可以用正态分布（的一点变形）来估计。仅一步的，结合 Theorem 5.5.17，在一定条件下我们可以用样本方差来代替方差。（见 Example 5.5.18）

p240 Example 5.5.19 这里问题是给定一个随机采样 $X_1, \ldots, X_n$，以及一个统计，在此为 $\hat{p}/(1-\hat{p})$，其中 $\hat{p} = \sum_i X_i/n$. 而我们希望做的就是得到这个统计的诸如方差等性质.

p242 (5.5.9) 注意这里的 θ 是各随机变量的期望组成的向量.

p242 Example 5.5.22 设 $g(x) = x/(1-x)$，注意 $E\hat{p} = p$，因此

$$\operatorname{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right) = \operatorname{Var}_p g(\hat{p}) \approx [g'(p)]^2 \operatorname{Var}_p \hat{p}.$$

p243 Example 5.5.25 TODO

p245 Sec 5.6 本节的目的可以理解为，在假设可以生成 uniform distributed 的随机数的情况下，考虑如何生成满足特定分布的随机数。具体的步骤分为两步，首先给出一个生成随机数的算法，接下来利用大数定理或者通过说明 cdf 相等来说明生成的随机数确实满足某个分布。

p251 Example 5.6.7 考虑如何估计一个不规则图形的面积：用一个正方形来框住该不规则图形，然后往正方形内随机的撒点，计算落在图形内的点的个数占总点数的比例，就可以大致估计出该图形在面积上占该正方形的比例。在这里的想法也是类似的。直观上来看，我们想做的事情是生成一些随机数，它们为某些值的概率符合某个分布，而这个方法里，对于 $f_X(x)$ 大的位置，它就更容易被接受。更数学的角度上看，5.6.11 表明生成的随机数确实满足所需分布。

p253 Theorem 5.6.8 考虑