第 6 讲*

特殊分布（离散型和连续型）

麻省理工学院 14.30 2006 年春季

Herman Bennett

15 离散型分布

我们已经学过二项分布和均匀分布。

15.1 超几何分布

设N个元素组成的总体，其中含有M个“成功”元素，每次抽样从中抽取n个元素中含有“成功”元素的个数为随机变量X。那么X的概率质量函数称为超几何分布，表达式为：

$$f(x)=P(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad \text{对于 } x=0,1,\ldots,n\text{。} \qquad (40)$$

相应的数学期望和方差为：

$$E(X)=\frac{nM}{N} \qquad \text{和} \qquad Var(X)=\left(\frac{N-n}{N-1}\right)\frac{nM}{N}\left(1-\frac{M}{N}\right)$$

注意：这些讲义不一定是自封的。它们只是对讲座的一种补充而不是替代。

15.2 负二项分布

二项分布计算的是在一个固定试验次数（n）中发生的成功次数。反之，我们也可计算给定成功次数（r）所需要进行的试验次数。

设随机变量X为得到r次“成功”所需要进行的试验次数。X的分布函数称为负二项分布，即：

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \text{对于} \quad x = r, r+1, r+2, \ldots \qquad \textbf{(41)}$$

相应的数学期望和方差为：

$$E(X) = \frac{r}{p} \qquad \text{和} \qquad Var(X) = \frac{r(1-p)}{p^2}$$

- $r = 1$ → 几何分布：“等待成功。”

15.3 泊松分布

随机变量X被称作参数为λ（$\lambda > 0$）的泊松分布，满足下式：

$$X \sim \mathcal{P}(\lambda)： \quad f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{对于} \quad x = 0, 1, 2, \ldots \qquad \textbf{(42)}$$

相应的数学期望和方差为：

$$E(X) = \lambda \qquad \text{和} \qquad Var(X) = \lambda$$

- λ表示一定单位的时间或一定单位的面积。

● 如果 X_1 和 X_2 是相互独立的随机变量，服从均值分别为 λ_1 和 λ_2 的泊松分布，那么存在随机变量 $Y = X_1 + X_2$，服从均值为 $\lambda_1 + \lambda_2$（随机变量的函数，讲义 5）的泊松分布。

● 说明：$\sum_{x=0}^{\infty} f(x) = e^{-\lambda} \underbrace{\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}}_{=e^{\lambda}} = 1$。

● 同前面两种分布一样，泊松分布不是在自然试验中得出的。

例 15.1. 设 X 服从泊松分布（λ），计算 $E(X)$。

例 15.2. 假设商场每天的客流量是服从参数为 λ 的泊松分布的随机变量。据了解，商场每天平均接待 20 位顾客，因此 $\lambda = 20$。计算以下情形的概率：（1）明天将有 20 位顾客（2）未来两天内将有 30 位顾客　（3）明天中午之前至少有 7 位顾客。

15.3.1 泊松分布和泊松过程

一个常见的混淆之处在于……

每单位时间内泊松比为 λ 的泊松过程是一个连续过程，它满足如下性质：

（1）在每个间隔长度为 t 的固定时间段内，发生的次数服从泊松分布，数学期望为 λt 。

（2）每两个不相连的时间间隔内发生的次数是相互独立的。

- 泊松过程：当试验包含 t 个时间单位时，使用 $\lambda\ t$ 。

例 15.3. 回答**例 15.2 中的问题**，其中，假定现在商场的客流量服从泊松过程（同样每天有平均 20 个顾客）。

- 泊松分布与二项分布。当 $n \to \infty$，$p \to 0$，且 $np \to \lambda$ 时，二项分布的极限 → 泊松分布。

16 连续型分布

我们已经学过均匀分布。

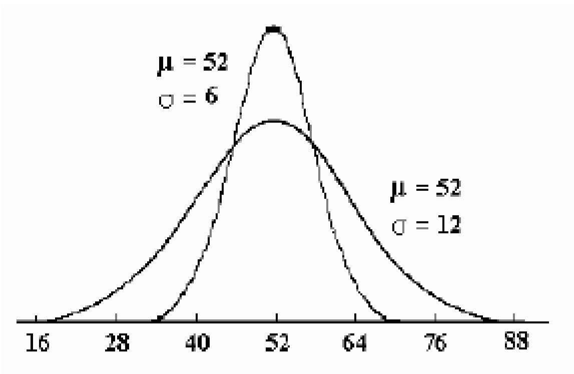
16.1 正态分布

一个随机变量X被称作参数为 μ 和 σ^2（$\sigma^2>0$）的正态分布，则X的概率密度函数表达式为：

$$X\sim N(\mu,\sigma^2)：\quad f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}\text{，对于 }-\infty<x<\infty \qquad (43)$$

相应的数学期望、方差和矩母函数分别为：

$$E(X)=\mu \qquad Var(X)=\sigma^2\text{，以及}\qquad E(e^{tX})=e^{t\mu+\frac{\sigma^2t^2}{2}}$$



- 为什么正态分布如此重要？

1. 正态分布的图形是大家非常熟悉的钟形。它提供了实际观测的理论基础，即许多随机现象服从（或至少近似服从）一种正态分布：

“任何偏离平均值越远的特定结果，发生的可能性越小；这一特征是以高于或低于均值而呈对称式分布。”

例：人口中每个人的身高和体重；物理量的测量误差；一个特定种子的蛋白质水平等。

2. 正态分布为其他分布提供了一个很好的近似值，比如泊松分布和二项分布。

3. 正态分布比其它钟状形态分布在分析上更易把握。

4. 中心极限定理（更多内容详见讲义 7）。

5. 正态分布对描绘人口分布（联系观点 1）有重要的帮助。

- 图形性质

1. 钟状形态而且呈对称状。

2. 在均值 μ 附近集中分布，与中位数重合。

3. 分散程度和平滑程度仅取决于方差 σ^2。

4. $P\left(\mu-\sigma < X < \mu+\sigma\right)=0.6826 \qquad \forall \mu,\sigma^2$!

5. $P\left(\mu-2\sigma < X < \mu+2\sigma\right)=0.9544 \qquad \forall \mu,\sigma^2$!

- 如果X～$N(\mu,\sigma^2)$，那么随机变量$Z=\left(X-\mu\right)/\sigma$满足Z～N（0，1）。$N\left(0,1\right)$分布称为<u>**标准正态**</u>分布，它的累积分布函数通常记作$F_Z\left(z\right)=\Phi\left(z\right)$。

- 正态分布的累积分布函数没有解析解，且它的值须在$N\left(0,1\right)$表中查找（见附录表）。

- 注意$\Phi(z)=1-\Phi(-z)$。事实上：$F_Y(y)=1-F_Y(-y)\quad \forall Y\sim N(0,\sigma^2)$。

- 如果$X_i\sim N(\mu_i,\sigma_i^2)$且所有的n个$X_i$相互独立，那么随机变量H表示为：

$$H=\sum_{i=0}^{n}\alpha_i X_i+b_i\sim N(\sum_{i=0}^{n}\alpha_i\mu_i+b_i,\sum_{i=0}^{n}\alpha_i^2\sigma_i^2)\tag{44}$$

例 16.1 用讲义 5 给出的工具，推导作为随机变量 $X\sim N(0,\sigma^2)$ 变换的$Z=(X-\mu)/\sigma$的分布。

例 16.2 计算$E(X)$，其中 $X\sim N(\mu,\sigma^2)$。

例 16.3 假设随机变量 X 服从均值为 5，标准差为 2 的正态分布。计算 $P(1<X<8)$ 和 $P(|X-5|<2)$。

例 16.4 假设有两种型号的灯泡（A 和 B）。A 的使用寿命服从均值为 100（小时），方差为 16 的正态分布。B 的使用寿命服从均值为 110（小时），方差为 30 的正态分布。（1）A 型号灯泡的使用寿命超过 110 小时的概率为多少？（2）如果灯泡 A 与灯泡 B 同时点亮，那么 A 持续的时间多于 B 的概率为多少？（3）两种型号灯泡的使用寿命都超过 105 小时的概率为多少？

- 二项分布可以近似地用正态分布表示。经验估计：$\min\left(np, n(1-p)\right) \geq 5$。

16.2 对数正态分布

如果X是随机变量，$\ln(X)$服从$N(\mu, \sigma^2)$，那么X是具有概率密度函数（随机变量变换）的对数正态分布，即

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{-\left(\ln(x)-\mu\right)^2/\left(2\sigma^2\right)}，对于\ 0 < x < \infty，\ -\infty < \mu < \infty，\ \sigma > 0 \quad \textbf{(45)}$$

$$\mathrm{Ln(X)} \sim \mathrm{N}(\mu, \sigma^2) \leftrightarrow \mathrm{X} \sim \mathrm{LnN}(\mu, \sigma^2)$$

相应的数学期望和方差为：

$$E(X) = e^{\mu+\left(\sigma^2/2\right)} \quad 和 \quad Var(X) = e^{2\left(\mu+\sigma^2\right)} - e^{2\mu+\sigma^2}。$$

- 如果 $\mathrm{X} \sim \mathrm{N}(\mu, \sigma^2)$，那么 $e^X \sim LnN(\mu, \sigma^2)$。

16．3 伽玛分布

随机变量X称为参数为α和β（$\alpha, \beta > 0$）的伽玛分布，则X的概率密度函数为：

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}， \quad 对于 0 < x < \infty \qquad \textbf{(46)}$$

这里，

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \ 是有限的，且\ \alpha > 0。$$

$\Gamma(\alpha) = (\alpha-1)!$如果$\alpha$是正整数，且$\Gamma(0.5) = \pi$。

相应的数学期望和方差为：

$$E(X)=\alpha\beta \qquad 和 \qquad Var(X)=\alpha\beta^2$$

- 假设一个泊松过程。设 Y 服从参数λ的泊松分布。定义 X 为等待第 r 个事件发生的时间。那么，X 为参数$\alpha=r$和$\beta=1/\lambda$的伽玛分布。

16.4 指数分布

随机变量X服从参数β（$\beta>0$）的指数分布， X的概密度率函数表达式为：

$$f(x)=\frac{1}{\beta}e^{-x/\beta}, \qquad 对于 0<x<\infty \tag{47}$$

相应的数学期望和方差为：

$$E(X)=\beta \qquad 和 \qquad Var(X)=\beta^2$$

- 指数分布是参数$\alpha=1$的伽玛分布。

16.5 χ^2分布

随机变量X称为参数$p>0$（自由度）的χ^2分布， X的概率密度函数为：

X～$\chi^2_{\{p\}}$： $f(x)=\dfrac{1}{\Gamma(p/2)2^{p/2}}x^{p/2-1}e^{-x/2}$，对于$0<x<\infty$和整数 p。 （48）

相应的数学期望和方差为：

$$E(X) = p \quad \text{和} \quad Var(X) = 2p$$

- χ^2 分布是参数 $\alpha = p/2$ 和 $\beta = 2$ 的伽玛分布。
- 如果 $Y \sim N(0,1)$，那么随机向量 $Z = Y^2$ 服从：

$$Z = Y^2 \sim \chi_1^2 \qquad \text{（随机变量变换）} \qquad \textbf{（49）}$$

- 如果 $X_1 \sim \chi^2_{(p)}$ 和 $X_2 \sim \chi^2_{(q)}$ 是相互独立的，那么随机变量 $H = X_1 + X_2$ 服从：

$$H = X_1 + X_2 \sim \chi^2_{(p+q\_} \qquad \text{（随机变量变换）} \qquad \textbf{（50）}$$

- 广泛应用于经济学。
- 单个分布与分布族（含有一个以上参数）的概念

16.6 二维正态分布

一个二维随机向量 (X_1, X_2) 称为一个二维正态分布，则 (X_1, X_2) 的概率密度函数表示为：

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-b/\left(2\left(1-\rho^2\right)\right)} \qquad \textbf{（51）}$$

$$\rho = Corr(X_1, X_2)$$

$$b \equiv \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}$$

- $\rho = 0 \leftrightarrow X_1$ 和 X_2 相互独立（仅在正态情况下）$\leftrightarrow f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$。