第十六章. 卡方分布.

卡方分布: 如果随机变量 X 服从自由度为 $\nu \geqslant 0$ 的卡方分布, 那么 X 的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{(\nu/2-1)}\mathrm{e}^{-x/2} & \text{若 } x\geqslant 0\\ 0 & \text{其他.}\end{cases}$$

我们将其记作 $X\sim\chi^2(\nu)$.

$$\int_0^{+\infty} f(x)\,dx$$

$$=\frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\int_0^{+\infty} x^{\frac{\nu}{2}-1}e^{-x/2}\,dx$$

$$\overset{t=\frac{x}{2}}{=\!=}\frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\int_0^{+\infty}(2t)^{\frac{\nu}{2}-1}e^{-t}\cdot 2\,dt$$

$$=\frac{2^{\nu/2}}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\int_0^{+\infty} t^{\frac{\nu}{2}-1}e^{-t}\,dt$$

$$=\frac{2^{\nu/2}\cdot\Gamma(\frac{\nu}{2})}{2^{\nu/2}\cdot\Gamma(\frac{\nu}{2})}$$

$$=1$$

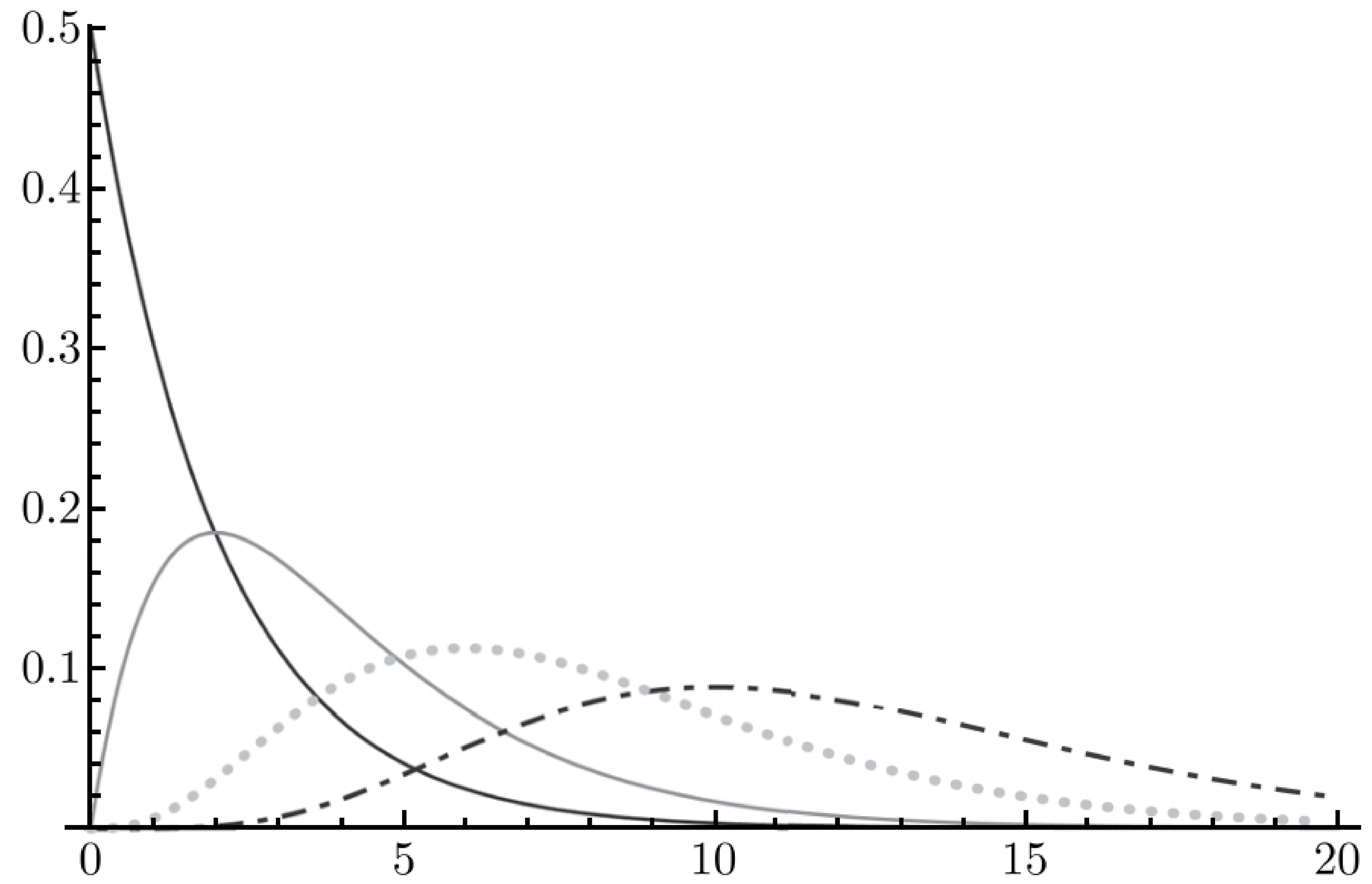

图 16-1 $\nu\in\{1,2,3,5,10,20\}$ 的卡方分布. 随着自由度的增加, 凸起的位置会向右移动

16.1 卡方分布的起源

卡方分布是由多个相互独立且服从正态分布的随机变量和构成的.

若 $X \sim N(0,1)$，则 $X^2 \sim \chi^2(1)$

标准正态分布　　　自由度为1的卡方分布

用累积分布函数证明：

$$F_{X^2}(y) = \text{Prob}(X^2 \le y)$$

$$= \text{Prob}(-\sqrt{y} \le X \le \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$\Rightarrow \quad f_{X^2}(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{2\sqrt{y}} \cdot [f_X(\sqrt{y}) + f_X(-\sqrt{y})]$$

$$= \frac{f_X(\sqrt{y})}{\sqrt{y}}$$

$$= \frac{y^{\frac{1}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{1}{2}} \cdot \Gamma(\frac{1}{2})}$$

得证

16.2 $X \sim \chi^2(1)$ 的均值与方差.

均值:

$$E(X) = \int_0^{+\infty} x \cdot \frac{1}{2^{\frac{1}{2}} \cdot \Gamma(\frac{1}{2})} x^{\frac{1}{2}-1} \cdot e^{-\frac{x}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^{\frac{1}{2}} e^{-\frac{x}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} (2t)^{\frac{1}{2}} e^{-t} \cdot 2dt$$

$$= \frac{2}{\sqrt{\pi}} \Gamma(\frac{3}{2})$$

$$= \frac{2}{\sqrt{\pi}} \cdot \frac{1}{2} \cdot \sqrt{\pi}$$

$$= 1$$

方差: 考虑 $\sigma^2 = E(X^2) - E^2(X)$

往求 $E(X^2)$

$$E(X^2) = \int_0^{+\infty} x^2 \cdot \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} x^{-\frac{1}{2}} e^{-\frac{x}{2}} dx$$

$$= \frac{1}{\sqrt{2}\,\Gamma(\frac{1}{2})} \int_0^{+\infty} x^{\frac{3}{2}} e^{-\frac{x}{2}} dx$$

$$= \frac{1}{\sqrt{2}\,\Gamma(\frac{1}{2})} \int_0^{+\infty} (2t)^{\frac{3}{2}} e^{-t} \cdot 2dt$$

$$= \frac{4}{\Gamma(\frac{1}{2})} \int_0^{+\infty} t^{\frac{3}{2}} e^{-t} dt$$

$$= \frac{4}{\Gamma(\frac{1}{2})} \cdot \Gamma(\frac{5}{2})$$

$$= 3$$

因此 $Var(X) = E(X^2) - E^2(X)$

$$= 3 - 1$$

$$= 2$$

16.3 卡方分布与服从正态分布的变量和.

卡方分布与服从正态分布的随机变量之和： 设 k 是一个正整数, $X_1, \cdots, X_k$ 是相互独立且均服从标准正态分布的随机变量, 这意味着 $X_i \sim N(0,1)$. 如果 $Y_k = X_1^2 + \cdots + X_k^2$, 那么 $Y_k \sim \chi^2(k)$. 更一般地, 设 $Y_{\nu_1}, \cdots, Y_{\nu_m}$ 是 m 个相互独立且均服从卡方分布的随机变量, 其中 $Y_{\nu_i} \sim \chi^2(\nu_i)$. 那么 $Y = Y_{\nu_1} + \cdots + Y_{\nu_m}$ 服从自由度为 $\nu_1 + \cdots + \nu_m$ 的卡方分布.

如果 $Y_{\nu_1} \sim \chi^2(\nu_1)$ 和 $Y_{\nu_2} \sim \chi^2(\nu_2)$ 是两个相互独立且均服从卡方分布的随机变量, 那么 $Y_{\nu_1} + Y_{\nu_2} \sim \chi^2(\nu_1 + \nu_2)$.

自由度为k（正整数）的卡方分布意味着k个服从正态分布的变量的平方和.

只需证明

$$\chi^2(\nu_1) + \chi^2(\nu_2) = \chi^2(\nu_1 + \nu_2)$$

即可. 其余可以用分组证明归纳补全.

16.3.1 直接积分求平方和.

$$\mathrm{Prob}(Y \le y) = \mathrm{Prob}(x_1^2 + x_2^2 \le y)$$

$$= \iint_{x_1^2 + x_2^2 \le y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} dx_1 dx_2$$

$$= \frac{1}{2\pi} \iint_{x_1^2 + x_2^2 \le y} e^{-\frac{x_1^2 + x_2^2}{2}} dx_1 dx_2$$

实际上是在一个半径为$\sqrt{y}$的圆盘上积分.

考虑极坐标变换法.

$$\begin{cases} X_1 = r\cdot\cos\theta. \\ X_2 = r\cdot\sin\theta. \end{cases}$$

$$原式 = \frac{1}{2\pi}\int_{r=0}^{\sqrt{y}}\int_{\theta=0}^{2\pi} e^{-\frac{r^2}{2}}\, r\,dr\,d\theta.$$

$$= \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\left(\int_{r=0}^{\sqrt{y}} e^{-\frac{r^2}{2}}\, r\,dr\right)d\theta$$

$$= \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\left[-e^{-\frac{r^2}{2}}\Big|_0^{\sqrt{y}}\right]d\theta.$$

$$= -1\times\frac{1}{2\pi}\int_{\theta=0}^{2\pi}\left(e^{-\frac{y}{2}}-1\right)d\theta.$$

$$= -1\times\frac{1}{2\pi}\left(e^{-\frac{y}{2}}-1\right)\cdot 2\pi$$

$$= 1-e^{-\frac{y}{2}}$$

$$= F_Y(y)$$

$$\Rightarrow f_Y(y) = \frac{d}{dy}F_Y(y)$$

$$= \frac{1}{2}\cdot e^{-\frac{y}{2}}$$

即自由度为2的卡方分布.

16.4 总结

卡方分布是一个稳定分布.

→ 泊松分布、正态分布

柯西分布.

为何有用？

举例. 工业误差常常服从正态分布.

但是误差和会被正负抵消. 加绝对值会导致函数不可积. 剩下的唯一方法便是变量平方和. 即卡方分布.

三大检验

- 卡方拟合优度检验.
- 卡方齐性检验
- 卡方独立性检验.