《天体物理中的统计方法》– 第 3 次作业 *

苏镇波†

中国科学技术大学物理学院天文系，合肥 230026

Gamma 分布的 mgf：

$$M(t) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \tag{1}$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^x (t-\lambda) dx \tag{2}$$

积分在 $t < \lambda$ 的时候会发散，因此可以写成：

$$M(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} (\frac{\Gamma(\alpha)}{(\lambda - t)^\alpha}) = (\frac{\lambda}{\lambda - t})^\alpha \tag{3}$$

求导后，可得：

$$M'(0) = E(X) = \frac{\alpha}{\lambda} \tag{4}$$

$$M''(0) = E(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2} \tag{5}$$

对于方差，有：

$$Var(X) = E(X^2) - [E(X)]^2 \tag{6}$$

$$= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} \tag{7}$$

$$= \frac{\alpha}{\lambda^2} \tag{8}$$

如果 X 和 Y 分别遵循参数为 α_1, λ、α_2, λ 得 gamma 分布，则 $X + Y$ 的 mgf 为：

$$(\frac{\lambda}{\lambda - t})^{\alpha_1} (\frac{\lambda}{\lambda - t})^{\alpha_2} = (\frac{\lambda}{\lambda - t})^{\alpha_1 + \alpha_2} \tag{9}$$

即具有相同 λ 值的独立 gamma 随机变量之和遵循伽马分布，因此具有 n 自由度的 chi-square 分布为 $\alpha = n/2, \lambda = \frac{1}{2}$ 的 gamma 分布，则有：

$$f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2} \tag{10}$$

其 mgf 为：

$$M(t) = (1 - 2t)^{-n/2} \tag{11}$$

*2022 春季《天体物理中的统计方法》

†Email: zbsu@mail.ustc.edu.cn, 学号：SA21022002

1 Problem 1

Answer.

如果 U,V 是互相独立的，且 $U \sim \chi_n^2, V \sim \chi_n^2$，那么根据式 (9)，$U+V$ 的 mgf 为:

$$(\frac{\lambda}{\lambda - t})^{n/2}(\frac{\lambda}{\lambda - t})^{m/2} = (\frac{\lambda}{\lambda - t})^{(m+n)/2} \tag{12}$$

即 $U+V \sim \chi_n^2$。

2 Problem 2

Answer.

根据式 (1-3) 的证明，对于 chi-square 而言，$\alpha = n/2$，$\lambda = \frac{1}{2}$，带入式 (3) 中，可得 χ_n^2 分布的 mgf 为：

$$M(t) = (1-2t)^{-n/2} \tag{13}$$

将 $\alpha = n/2, \lambda = \frac{1}{2}$ 带入式 (4) 和式 (8)，可得：平均数为 n，方差为 $2n$。

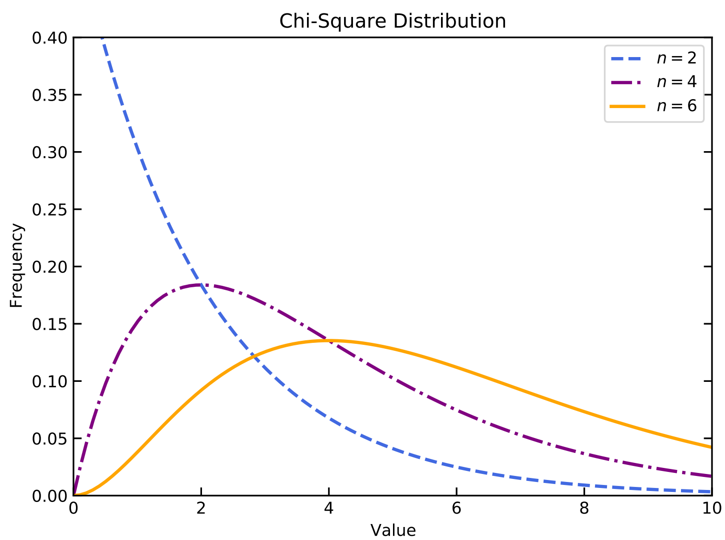

图 1: 如图，蓝色、紫色和橙色分别对应 Chi-square 下，n=2,4,6 的情况。

3 Problem 3

Answer.

根据 Taylor series expansion 到 2^{nd}，可得：

$$Y = g(X) \approx g(\mu_X) + (X - \mu_X)g'(\mu_X) + \frac{1}{2}(X - \mu_X)^2 g''(\mu_X) \tag{14}$$

因此对于 $Y = \mathrm{lnX}$ 有：

$$g'(\mu_X) = \frac{1}{\mu_X} g''(\mu_X) = -\frac{1}{\mu_X} \tag{15}$$

所以：

$$\mu_Y \approx \mathrm{ln}\mu_{\mathrm{X}} - \frac{\sigma_{\mathrm{X}}^2}{\mu_{\mathrm{X}}^2} \tag{16}$$

$$\sigma_Y^2 \approx \sigma_X^2 \times (\frac{1}{\mu_X^2}) = \frac{\sigma_X^2}{\mu_X^2} \tag{17}$$

4 Problem 4

Answer.

考虑 $Y = X^{\frac{1}{3}}$，X 在区间 [0, 1] 上，有:

精确解：

$$E(Y) = \int_0^1 x^{1/3} dx = \frac{3}{4} \tag{18}$$

$$E(Y^2) = \int_0^1 x^{2/3} dx = \frac{3}{5} \tag{19}$$

$$Var(Y) = E(Y^2) - E^2(Y) = \frac{3}{5} - \frac{3}{4}^2 = 0.0375 \tag{20}$$

$$\sigma_Y = \sqrt{Var(Y)} \approx 0.194 \tag{21}$$

$\mu_X = \frac{1}{2}$，有：

$$g'(X) = \frac{1}{3} x^{-2/3} \tag{22}$$

$$g''(X) = -\frac{2}{9} x^{-5/3} \tag{23}$$

$$Var(X) = \frac{1}{12} \tag{24}$$

$$g'(\mu_X) = \frac{2^{2/3}}{3} \approx 0.53 \tag{25}$$

$$g''(\mu_X) = \frac{2^{8/3}}{9} \approx -0.71 \tag{26}$$

所以有近似解：

$$E(Y) \approx 2^{-1/3} + \frac{1}{2} \times (\frac{1}{12} \times \frac{2^{8/3}}{9}) \approx 0.7643 \tag{27}$$

$$Var(Y) = 0.0233 \tag{28}$$

$$\sigma_Y = \sqrt{\frac{1}{12}} \times \frac{2^{2/3}}{3} \approx 0.15274 \tag{29}$$

同理，X 在区间 [1, 2] 上，有精确解：

$$E(Y) = \int_1^2 x^{1/3} dx \approx 1.13988 \tag{30}$$

$$E(Y^2) = \int_1^2 x^{2/3} dx \approx 1.3049 \tag{31}$$

$$Var(Y) = E(Y^2) - E^2(Y) = 1.3049 - 1.13988^2 \approx 0.0056 \tag{32}$$

$$\sigma_Y = \sqrt{Var(Y)} \approx 0.075 \tag{33}$$

$\mu_X = \frac{3}{2}$，有：

$$g'(\mu_X) \approx 0.254 \tag{34}$$

$$g''(\mu_X) \approx -0.113 \tag{35}$$

近似解：

$$E(Y) \approx \frac{3}{2}^{1/3} + \frac{1}{2} \times (\frac{1}{12} \times -0.113) \approx 1.14 \tag{36}$$

$$Var(Y) = \frac{0.254^2}{12} \approx 0.00539 \tag{37}$$

$$\sigma_Y = \sqrt{\frac{1}{12}} \times 0.254 \approx 0.073433 \tag{38}$$

对于区间 [0,1]，均值 E(Y) 精确解和近似解分别为：0.75，0.7643，偏离 1.9%；σ_Y 精确解和近似解分别为：0.194，0.15274，偏离 21.2%；

对于区间 [1,2]，均值 E(Y) 精确解和近似解分别为：1.13988，1.14，偏离 0.01%；σ_Y 精确解和近似解分别为：0.075，0.073433，偏离 2.1%；

在区间 [1,2] 上表现更好。

5 Problem 5

Answer.

根据高斯分布，利用 MCMC 方法有：

$$\hat{I}(f) = \frac{1}{N}(\frac{1}{\sqrt{2\pi}})\sum_{i=1}^{N} e^{-X_i^2/2} \tag{39}$$

代码形式为：

```
import numpy as np
import random

def mcmc(N, x_range):
    """
    N: Numbers of MCMC method iteration times.
    x_range: integration limits
    """
    sum = 0
    for i in range(N):
        x = random.uniform(0, int(x_range))
        sum += np.exp(-x**2/2)

    return 1/N * (1/np.sqrt(2*np.pi)) * sum

print(N = mcmc(10000, x_range = 1)) #0.34174123263753037

print(N = mcmc(10000, x_range = 5)) #0.4940471635
```

当积分区间为 [0, 1] 时，结果为 0.3417；

当积分区间为 [0, 5] 时，结果为 0.4940，接近 $\frac{1}{2}$。

6 Problem 6

Answer.

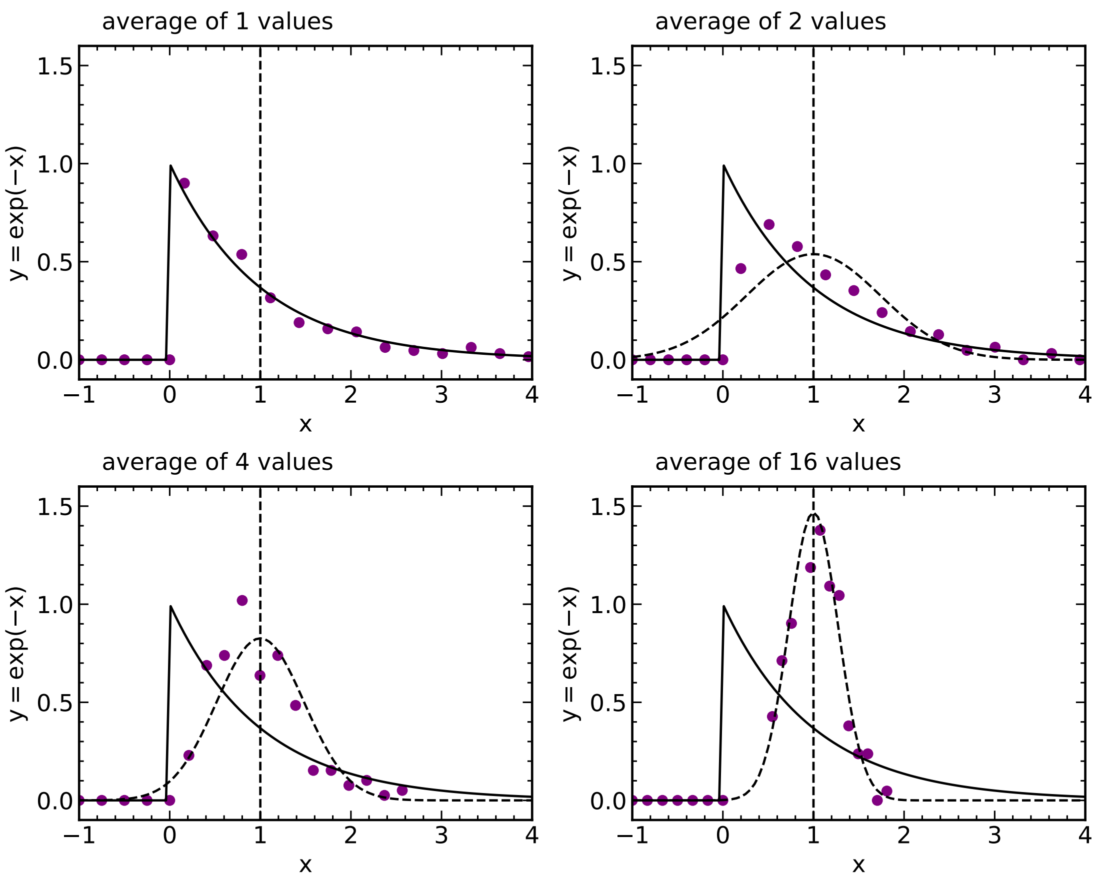

图 2: 上图中，紫色原点为随机生成的指数分布，黑色实线为初始指数分布，黑色虚线为 overplot 的最佳拟合高斯轮廓。分别为平均 1、2、4、16 个，生成随机数为 200。