

题 1

1)

由题意 $X_1, X_2, \dots, X_N \sim \text{Poisson}(\lambda)$

即 $Pr(X = k) = \frac{\lambda^k}{k!} * e^{-\lambda}$

设给定样本 $x_1 x_2 \dots x_N$

则似然函数 Likelihood Function

$$L(\lambda | x_1 x_2 \dots x_N) = \prod_{i=1}^N Pr(X = x_i)$$

log likelihood function

$$\begin{aligned} l(\lambda | x_1 x_2 \dots x_N) &= \ln[L(\lambda | x_1 x_2 \dots x_N)] \\ &= \sum_{i=1}^N \ln[Pr(X = x_i)] \\ &= \sum_{i=1}^N \ln\left[\frac{\lambda^{x_i}}{x_i!} * e^{-\lambda}\right] \\ &= \sum_{i=1}^N [x_i \ln(\lambda) + \ln(e^{-\lambda}) - \ln(x_i!)] \\ &= (x_1 x_2 \dots x_N) \ln(\lambda) - N\lambda - \sum_{i=1}^N \ln(x_i!) \\ \frac{\partial(l(\lambda | x_1 x_2 \dots x_N))}{\partial \lambda} &= \frac{x_1 x_2 \dots x_N}{\lambda} - N = 0 \\ \Rightarrow \hat{\lambda} &= \frac{x_1 x_2 \dots x_N}{N} = \bar{X} \end{aligned}$$

即当 $\lambda = \hat{\lambda} = \bar{X}$ 时, log likelihood function 与 likelihood function 取最大值, 即 $\hat{\lambda}$ 是 λ 的 MLE 估计.

由 MLE 的参数不变性, $\hat{\lambda}$ 是 λ 的 MLE 估计, 则 $\hat{\theta} = f(\hat{\lambda}) = e^{\bar{X}}$ 是 $\theta = f(\lambda) = e^{\lambda}$ 的 MLE 估计.

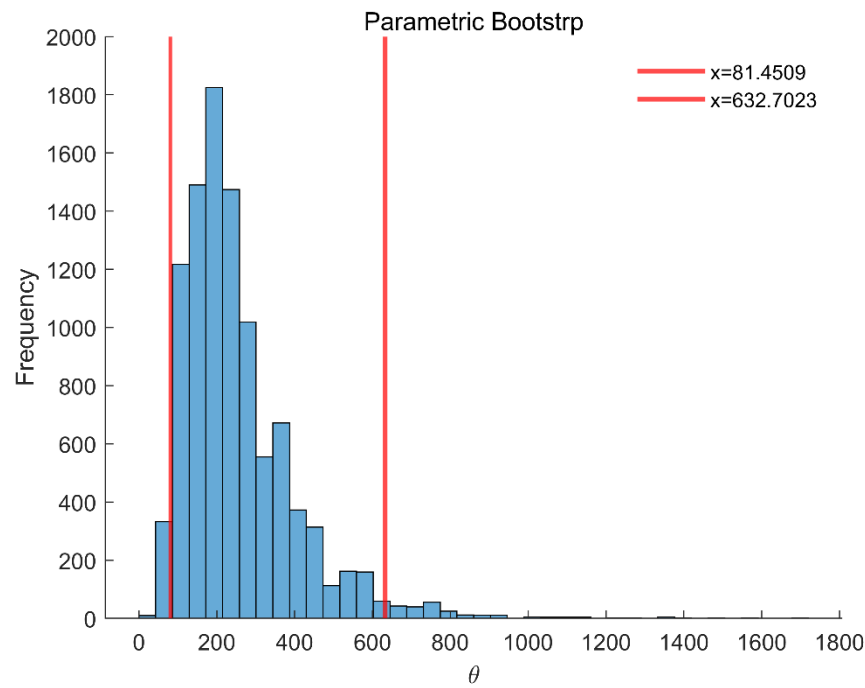
2)3)4)

由 parametric bootstrap(重采样 10000 次)得到的 $\hat{\theta}$ 的 95% 置信区间为 [81.45, 632.70], 见图 1-1。

由 Non-parametric bootstrap(重采样 10000 次)得到的 $\hat{\theta}$ 的 95% 置信区间为

[85.63, 601.845] ，见图 1-2。

即由真实的 θ (采样 10000 次)得到的 $\hat{\theta}$ 的 95%置信区间为[57.40, 403.43]，见图 1-3。



注：i) 图中红线表示 $\hat{\theta}$ 的 95%置信区间(下同)

图 1-1 由 parametric bootstrap 得到的 $\hat{\theta}$ 的分布

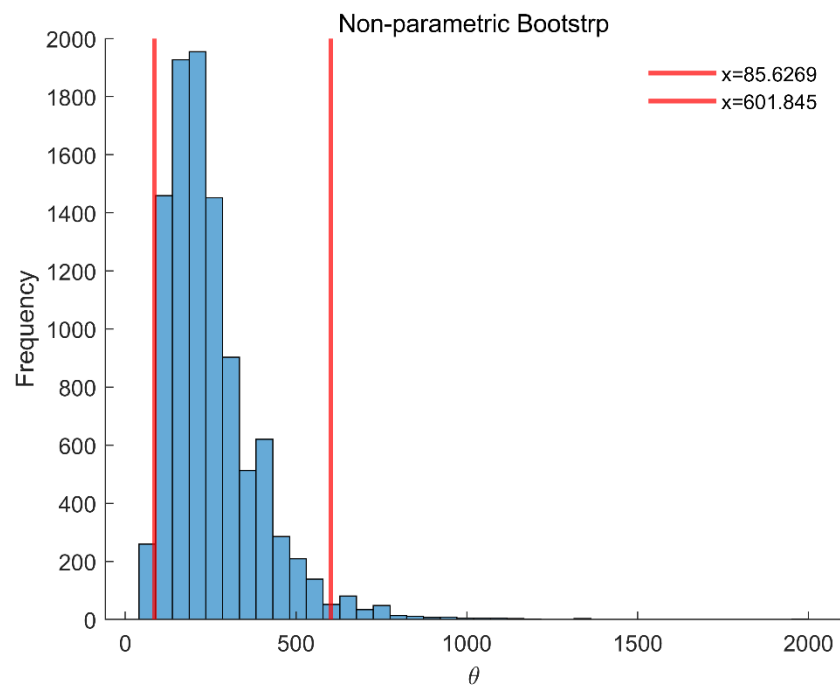


图 1-2 由 Non-parametric bootstrap 得到的 $\hat{\theta}$ 的分布

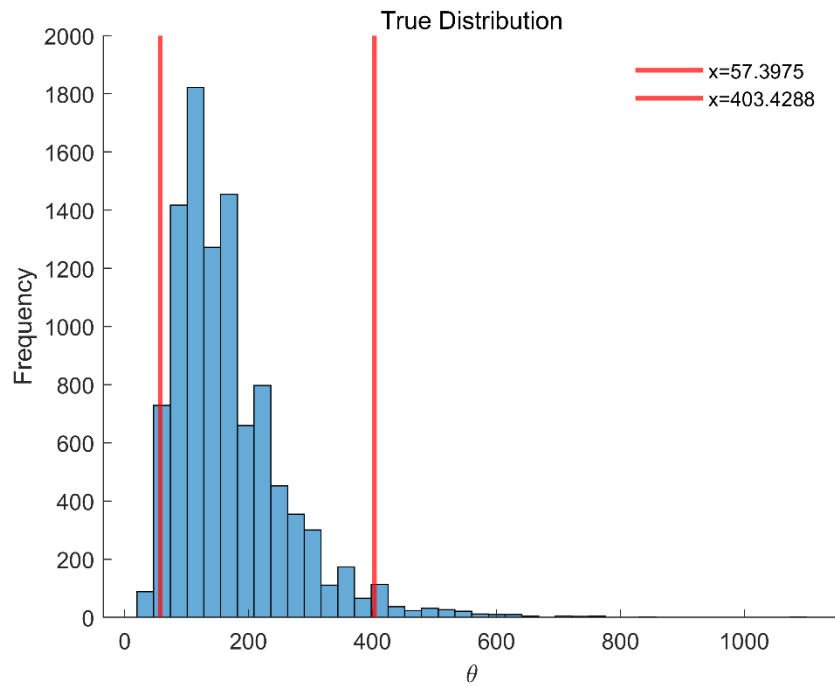


图 1-3 由真实的 θ 得到的 $\hat{\theta}$ 的分布

题 2

1)

由模型 $C_i \sim b(1, f(I))$, $f(I) = \frac{1}{1+e^{-\beta(x-\alpha)}}$, $\alpha = 45$, $\beta = 0.1$, 心理物理学曲线及相应的模拟数据见图 2-1。

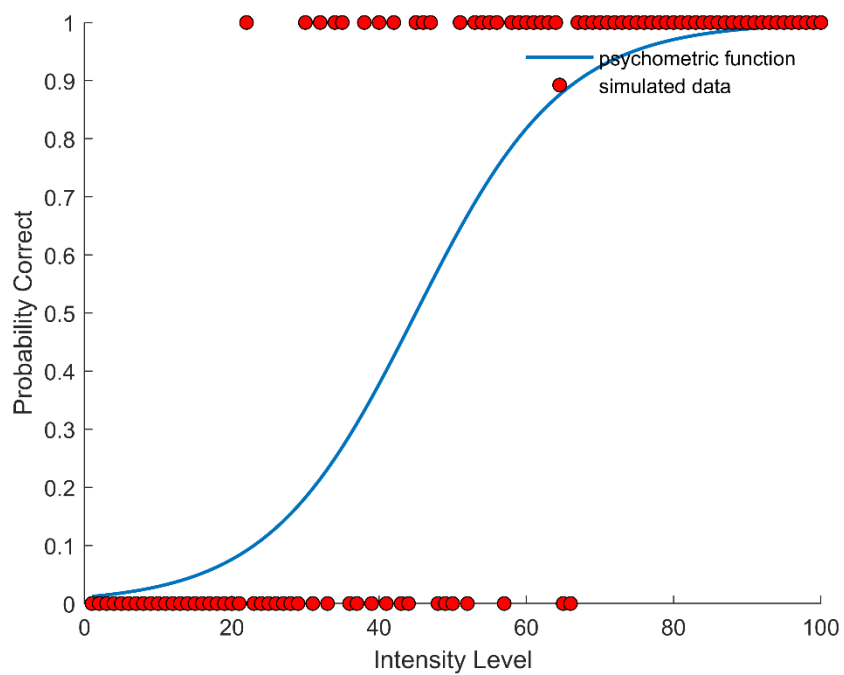


图 2-1 心理物理学曲线及模拟数据点

2)

α 的 MLE 估计为 43.39, β 的 MLE 估计为 0.10, 估计的心理物理学曲线见图 2-2.

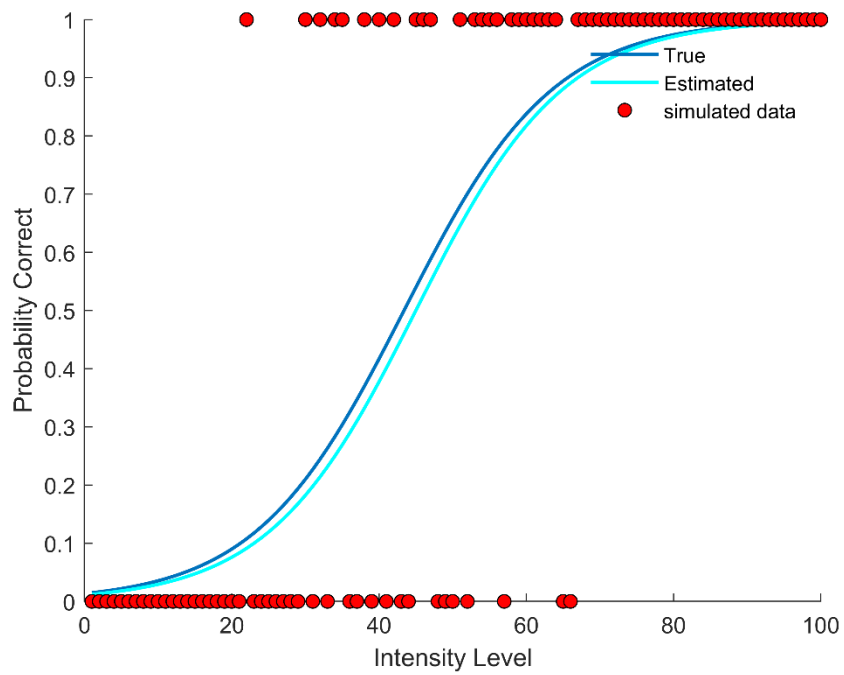


图 2-2 MLE 估计的心理物理学曲线

3)

由 parametric bootstrap(重采样 1000 次)得到的 $\hat{\alpha}$ 的 95%置信区间为[37.23, 49.49], $\hat{\beta}$ 的 95%置信区间为[0.07, 0.17].

题 3

1)

Steven’s model 和 Fechner’s model 及模拟数据见图 3-1.

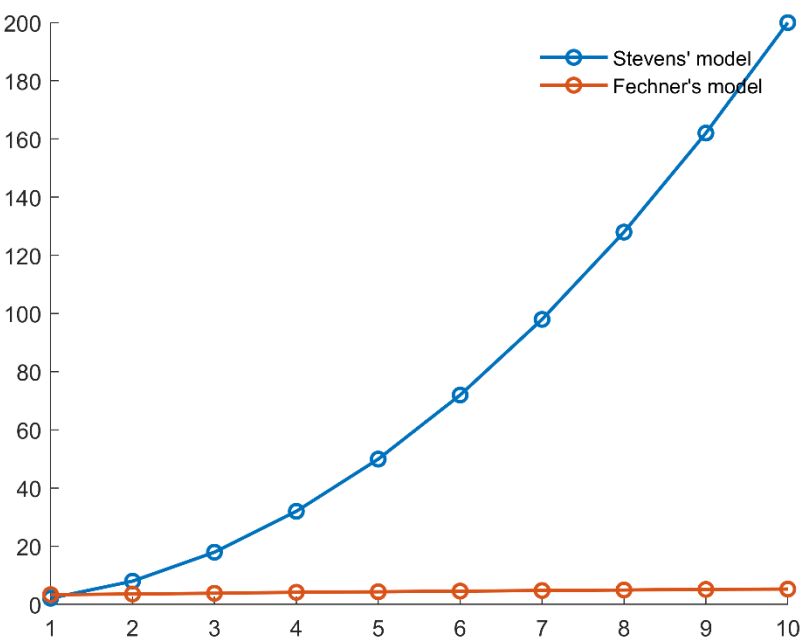


图 3-1 Stevens’ model 和 Fechner’s model 及模拟数据

2)

利用 MLE 估计和 BIC 标准，Stevens’ Model 与 Fechner’s Model 的模型比较见表 1

表 1. Stevens’ Model 与 Fechner’s Model 的模型比较						
数据来源	Mean -lnL		Mean BIC		Percentage of best fit	
	Stevens’	Fechner’s	Stevens’	Fechner’s	Stevens’	Fechner’s
Stevens’ Model	3.17	50.47	13.24	107.85	1.00	0.00
Fechner’s Model	3.46	3.36	13.83	13.62	0.50	0.50

3)

利用 MLE 估计和 LOOCV, Stevens' Model 与 Fechner's Model 的模型比较见表 2

表 2. Stevens' Model 与 Fechner's Model 的模型比较

数据来源	Mean CV		Percentage of best fit	
	Stevens'	Fechner's	Stevens'	Fechner's
Stevens' Model	9.32	53.66	1.00	0.00
Fechner's Model	9.73	9.67	0.42	0.58

4)

对任一数据集, 有 $CV = \sum_{j=1}^N -\ln L(y_j|\hat{\theta}_j) \geq -\ln L(y|\hat{\theta})$

证:

设 Leave-on-out crossvalidation 中, train 后得到的参数为 θ_j , 作为 test 的是 data(j)

设 MLE 中, 得到的参数为 θ

由数据集中所有数据点的最大似然估计可得

$$\begin{aligned} & -\ln[L(y|\hat{\theta})] \\ &= \sum_{i=1}^N -\ln P(y|\hat{\theta}) \leq \sum_{i=1}^N -\ln P(y|\hat{\theta}_j) \end{aligned}$$

由 train 时数据点的最大似然估计可得

$$\sum_{i \neq j} -\ln P(y|\hat{\theta}_j) \leq \sum_{i \neq j} -\ln P(y|\hat{\theta})$$

即

$$\sum_{i \neq j} \ln P(y|\hat{\theta}) \leq \sum_{i \neq j} \ln P(y|\hat{\theta}_j)$$

上下相加, 有

$$-\ln P(y_j|\hat{\theta}) \leq -\ln P(y_j|\hat{\theta}_j)$$

故而

$$-\ln L(y|\hat{\theta}) = \sum_{j=1}^N -\ln P(y_j|\hat{\theta}) \leq \sum_{j=1}^N -\ln P(y_j|\hat{\theta}_j) = CV$$