第四章 最大似然参数估计

概率基础

概率的定义

给事件 a 分配概率 $P(a)$，反映我们对事件 a 的发生的期望。

概率的特性

1. 事件的概率必须介于 0 和 1 之间
2. 所有可能结果的概率之和必须等于 1
3. 对于互斥事件，多个事件其一发生的概率，等于所有事件单独发生的概率之和

联合概率： $P(a, b)$

条件概率： $P(a \mid b)$

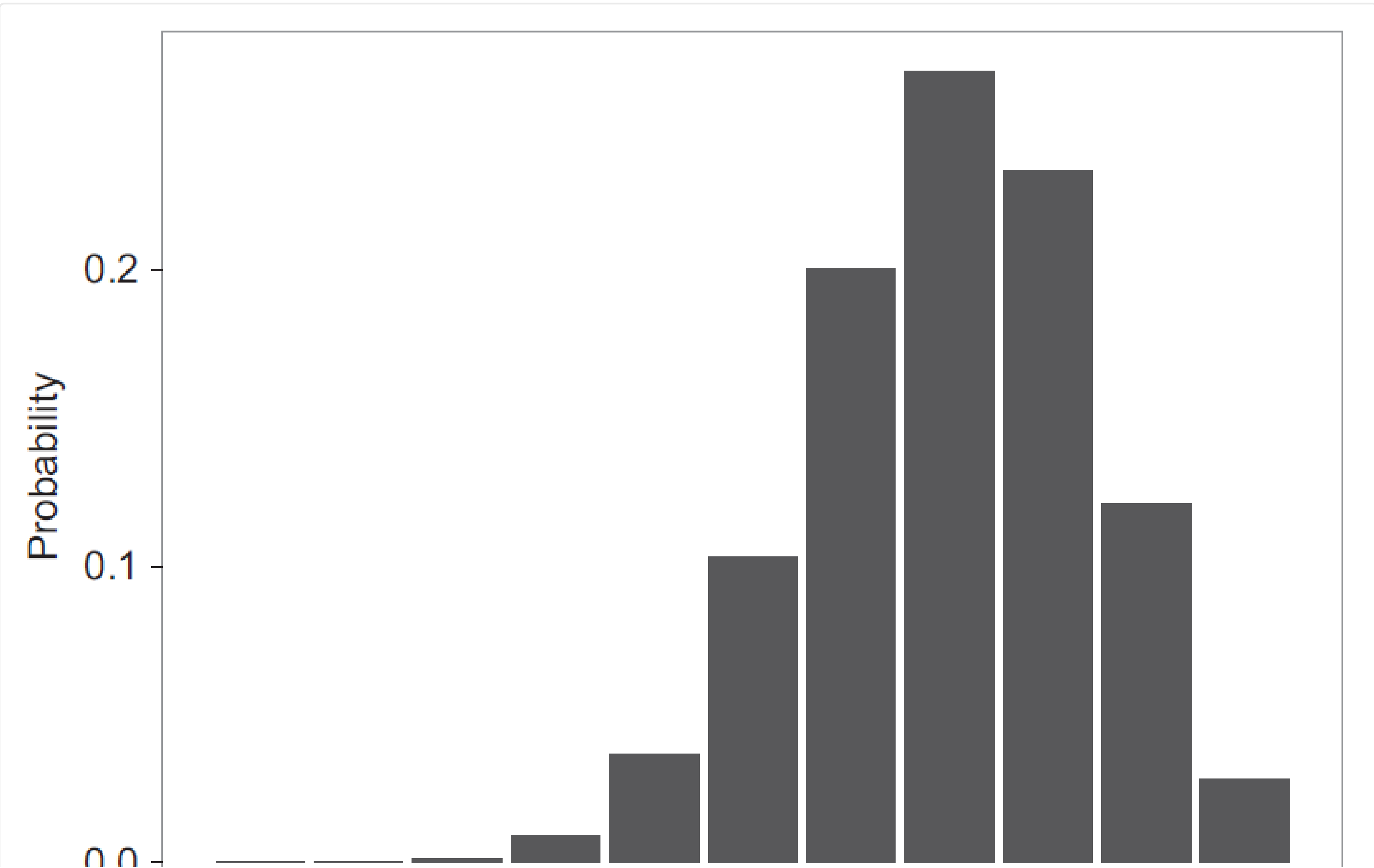
用 $P(data \mid model)$ 评估模型对数据进行预测的能力。

概率函数

离散事件：**概率质量函数**。

实例：GCM 计算对象与存储样本相似度，然后使用 Luce 选择规则分类，在只有两个类别 A 和 B 的情况下，在给定刺激 i 的条件下，归类为 A 的概率表示为 $P(R_i = A \mid i)$，简单地写作 P_A 。

假设我们对刺激 i 进行 10 次测试，多少次它归类为 A？ 下图中 $P_A = 0.7$



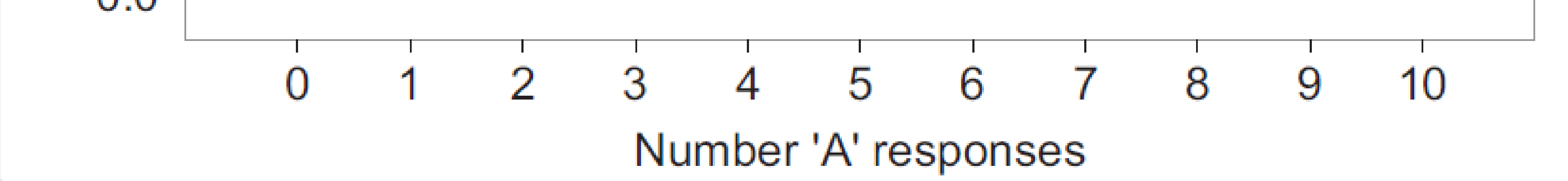


图 1 概率质量函数（PMF）范例

两种方式表示连续变量的概率分布：

累积分布函数，CDF

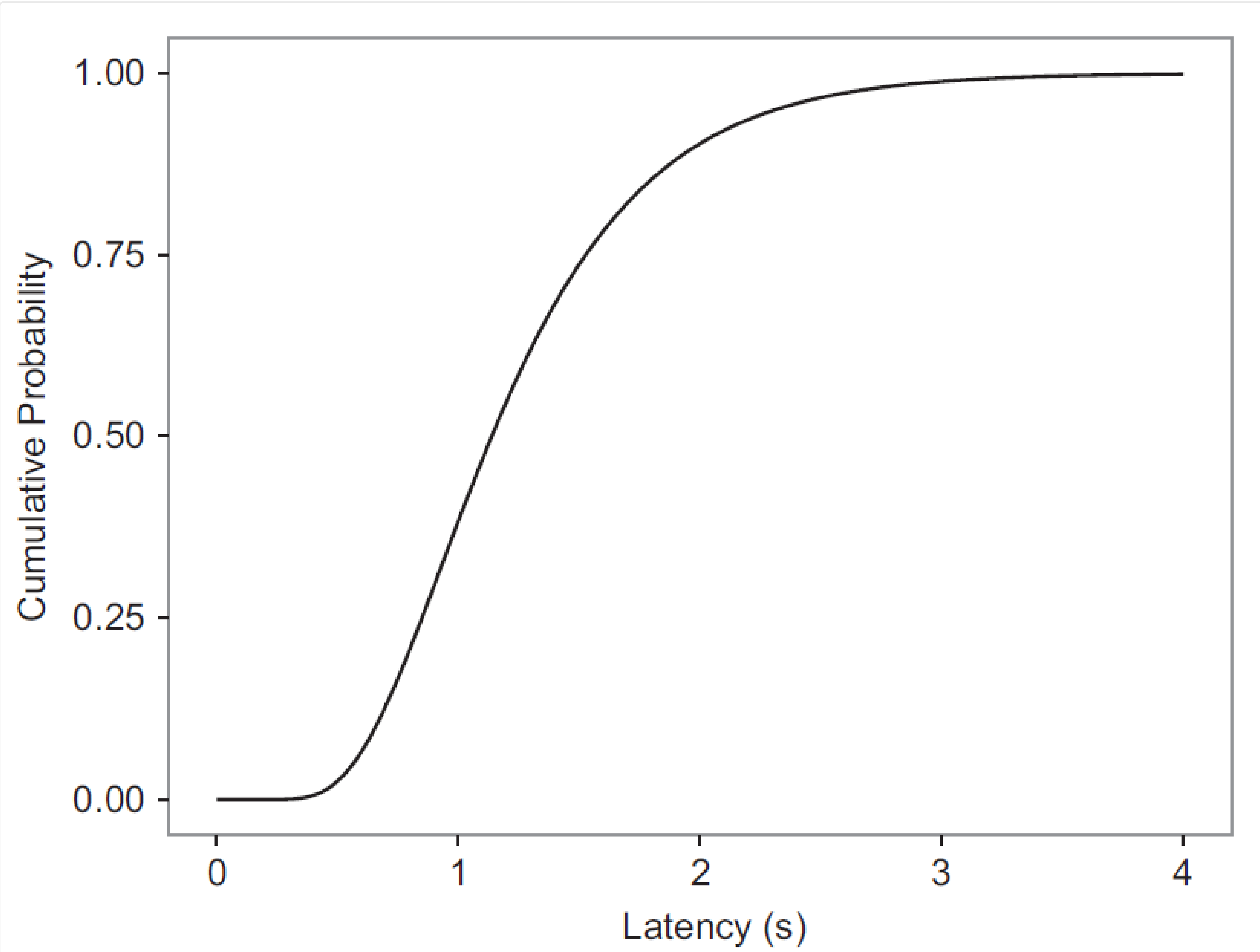


图 2 累计分布函数（CDF）示例

概率密度函数，PDF

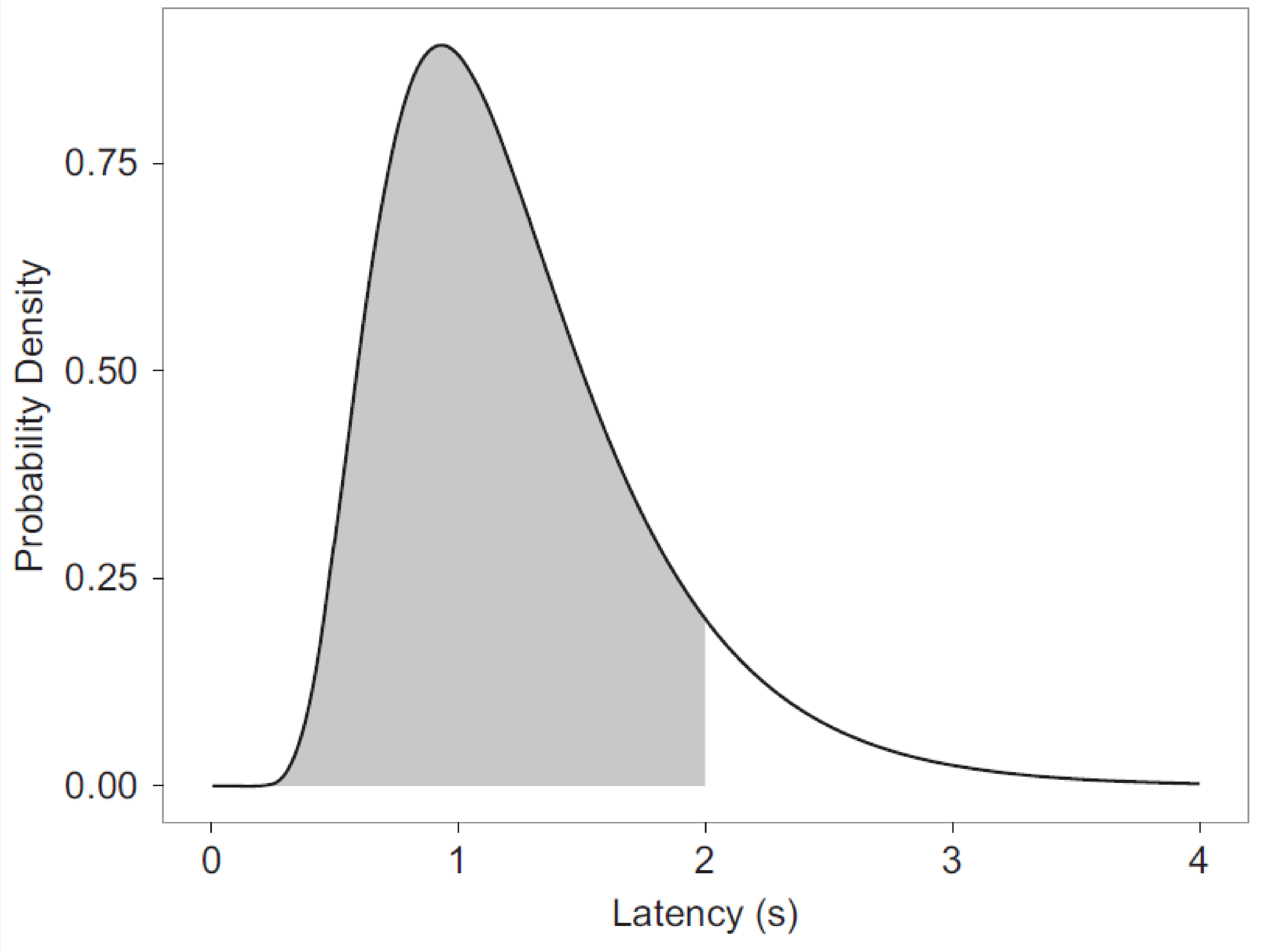


图 3 概率密度函数（PDF）示例

什么是似然

对于单个数据点 y ，模型 M 和参数值向量 θ ，在模型和参数给定情况下，得到此观测点的概率表示为 $f(y \mid \theta, M)$ 。

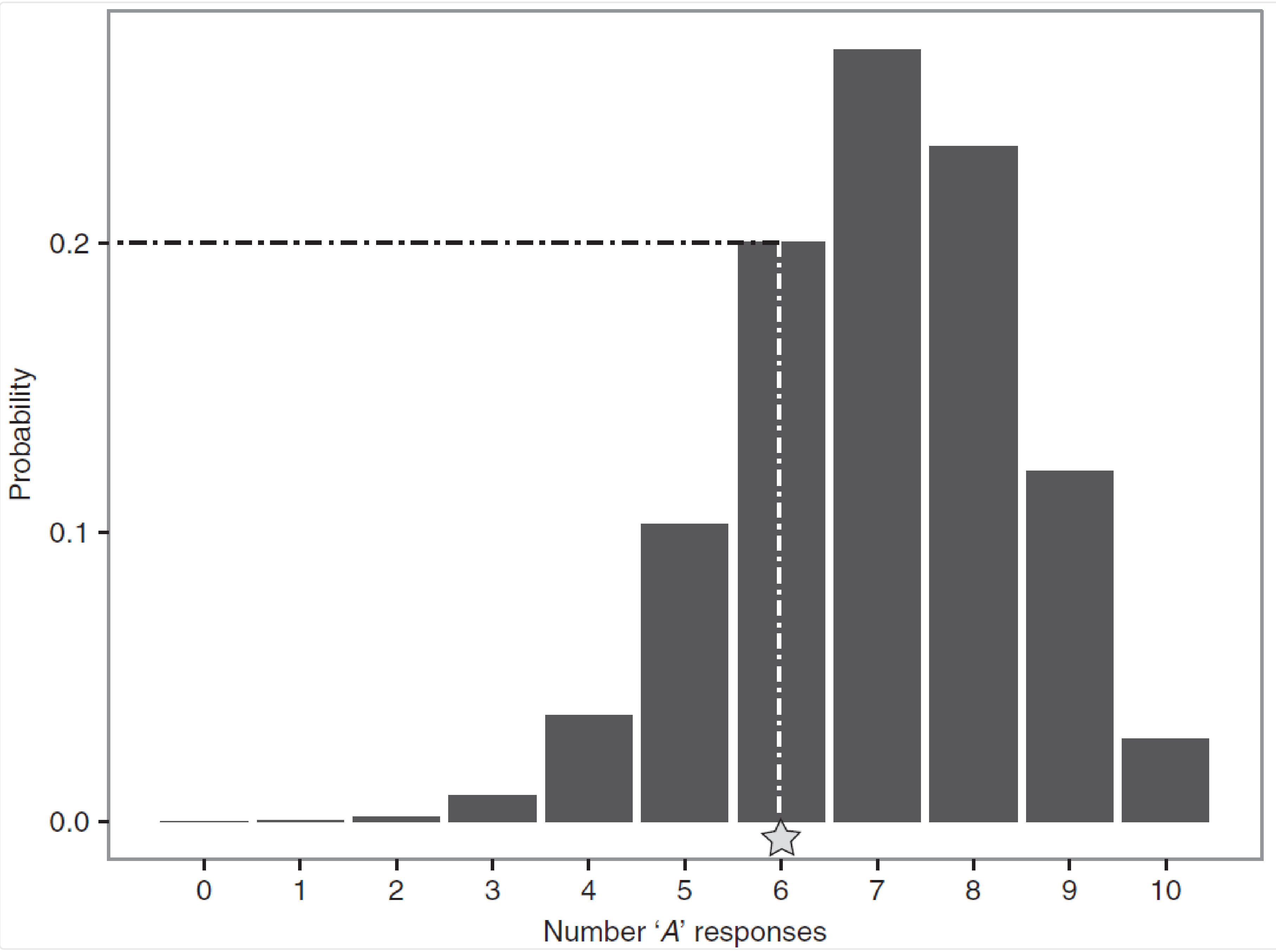


图 4 读取离散数据的概率

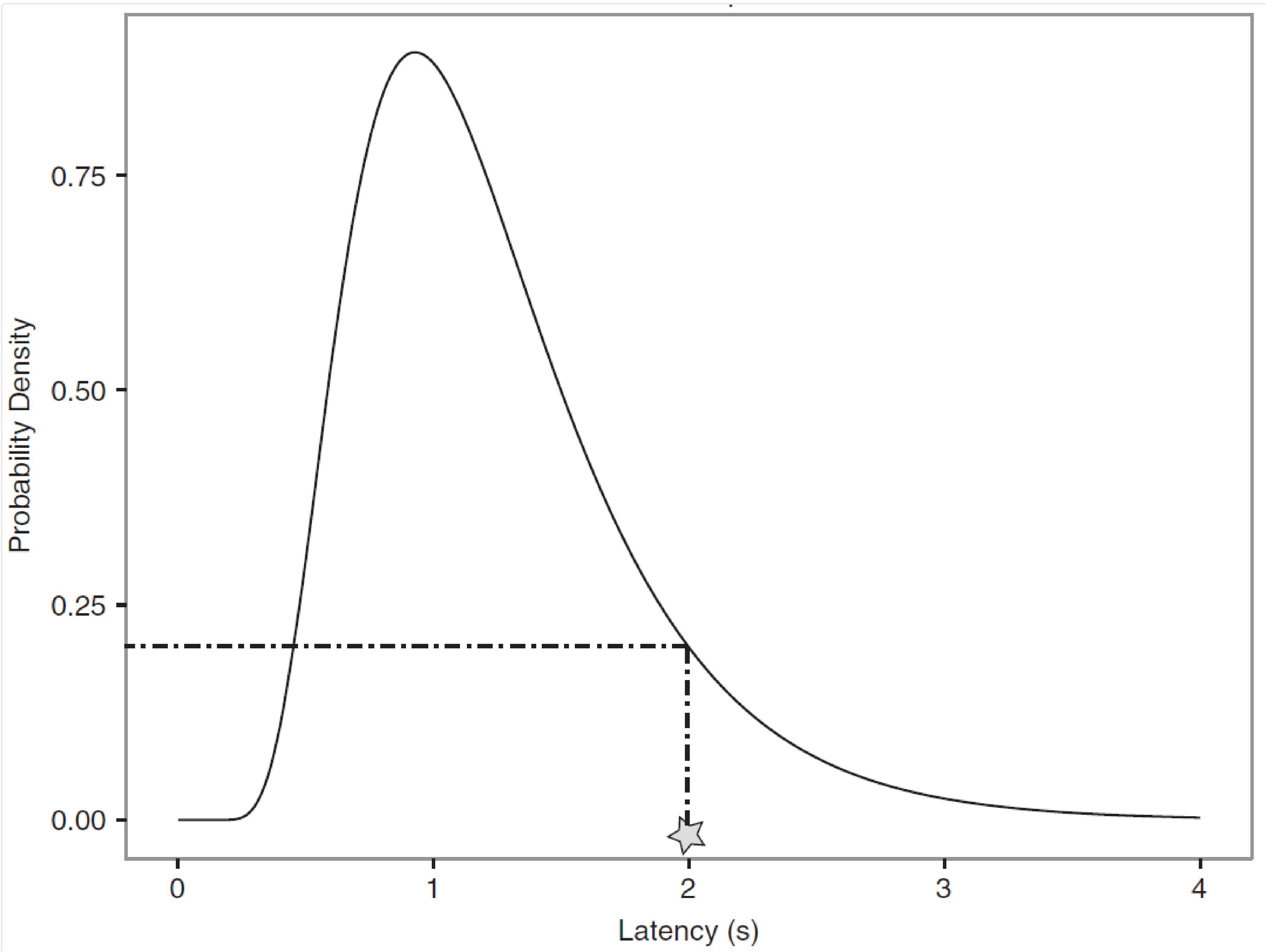


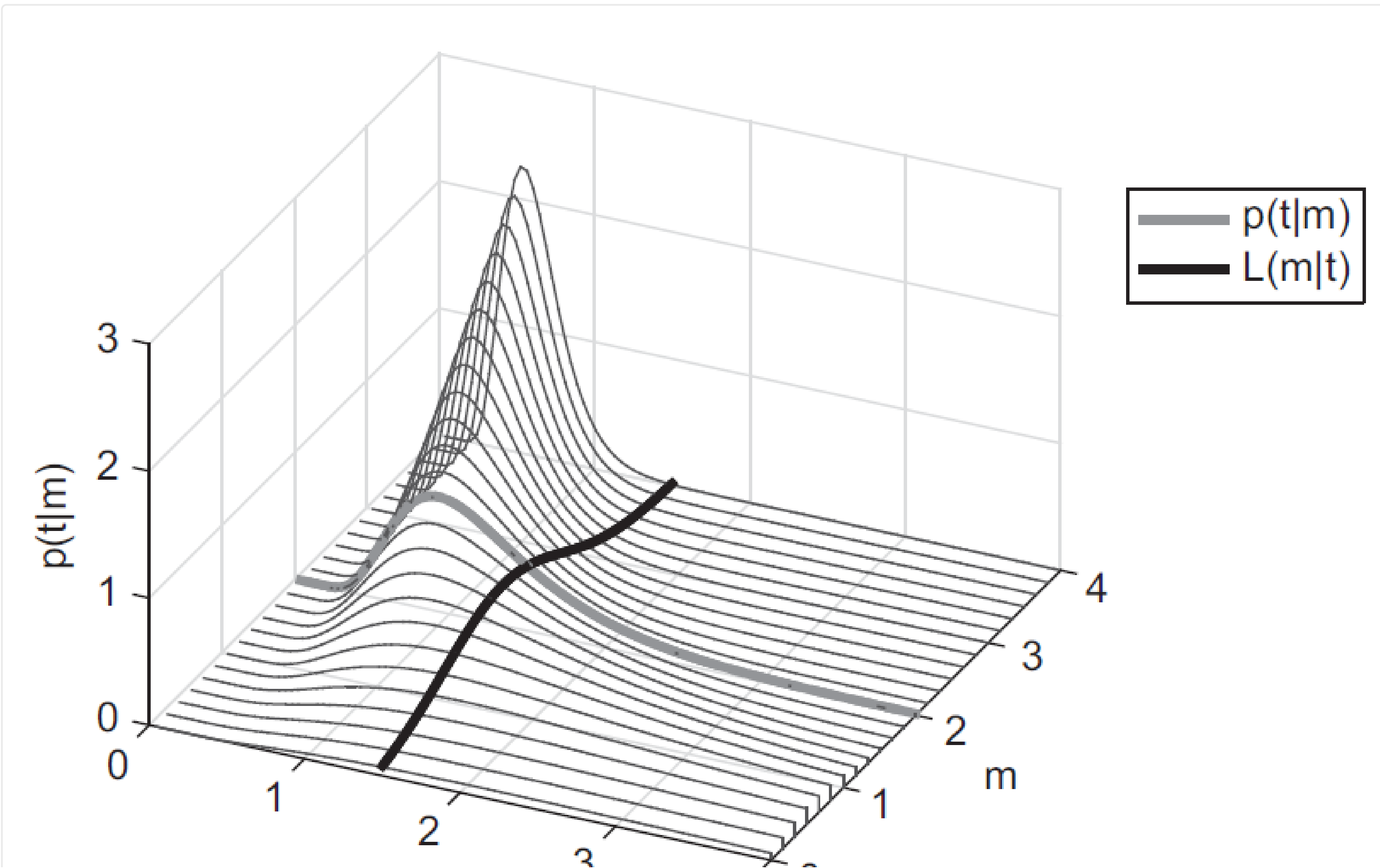
图 5 读取连续数据的概率

若是获得出现一个观测集合的概率，假设集合间观测值都相互独立，则可用：

$$f(\mathbf{y} \mid \boldsymbol{\theta}) = \prod^{k} f\left(y_k \mid \boldsymbol{\theta}\right)$$

概率函数告诉我们，在给定参数向量 θ 的情况下，数据 y 的概率。

似然函数告诉我们，在给定数据 y 的条件下，参数向量 θ 的似然。



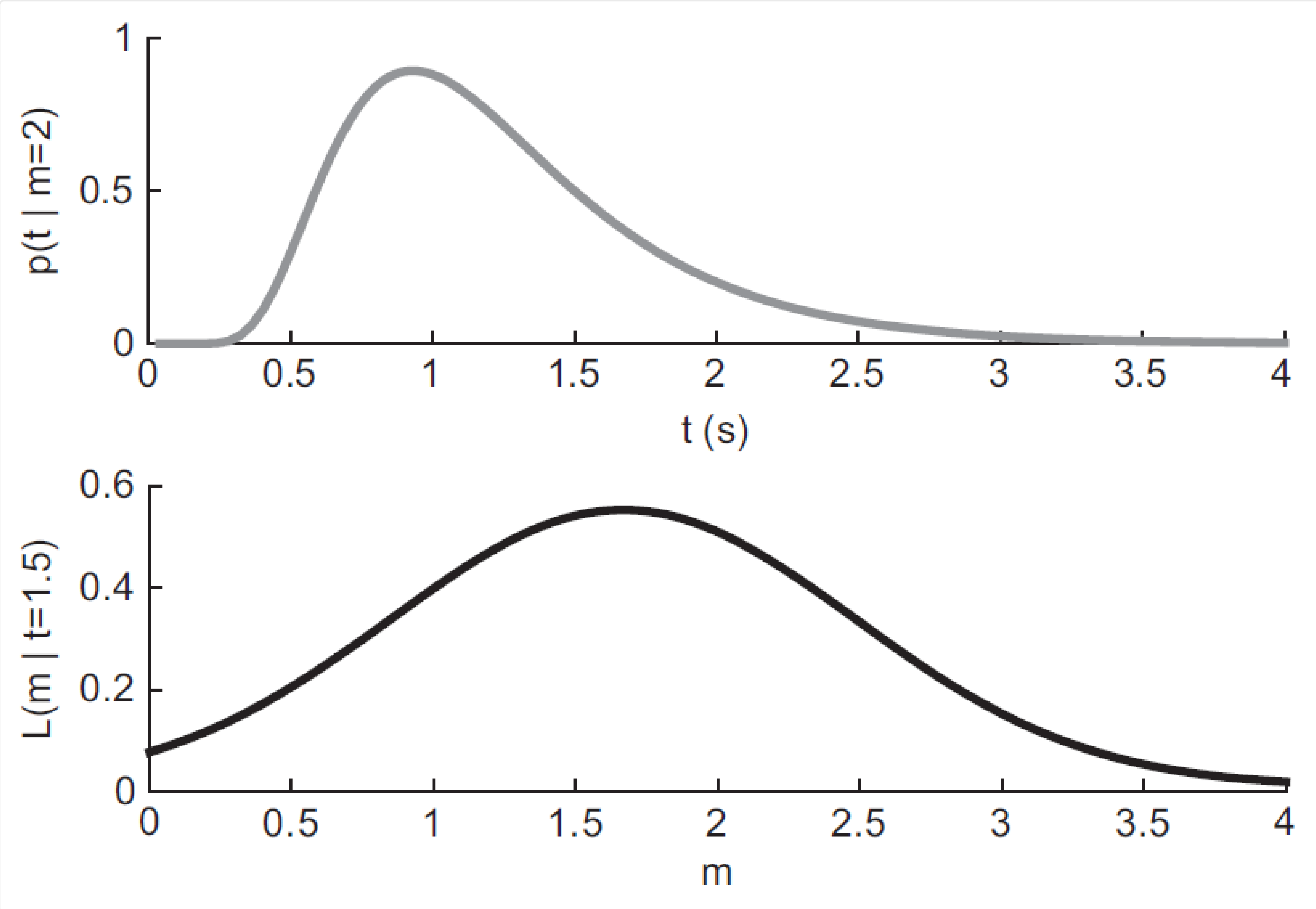


图 6 概率与似然的直观图

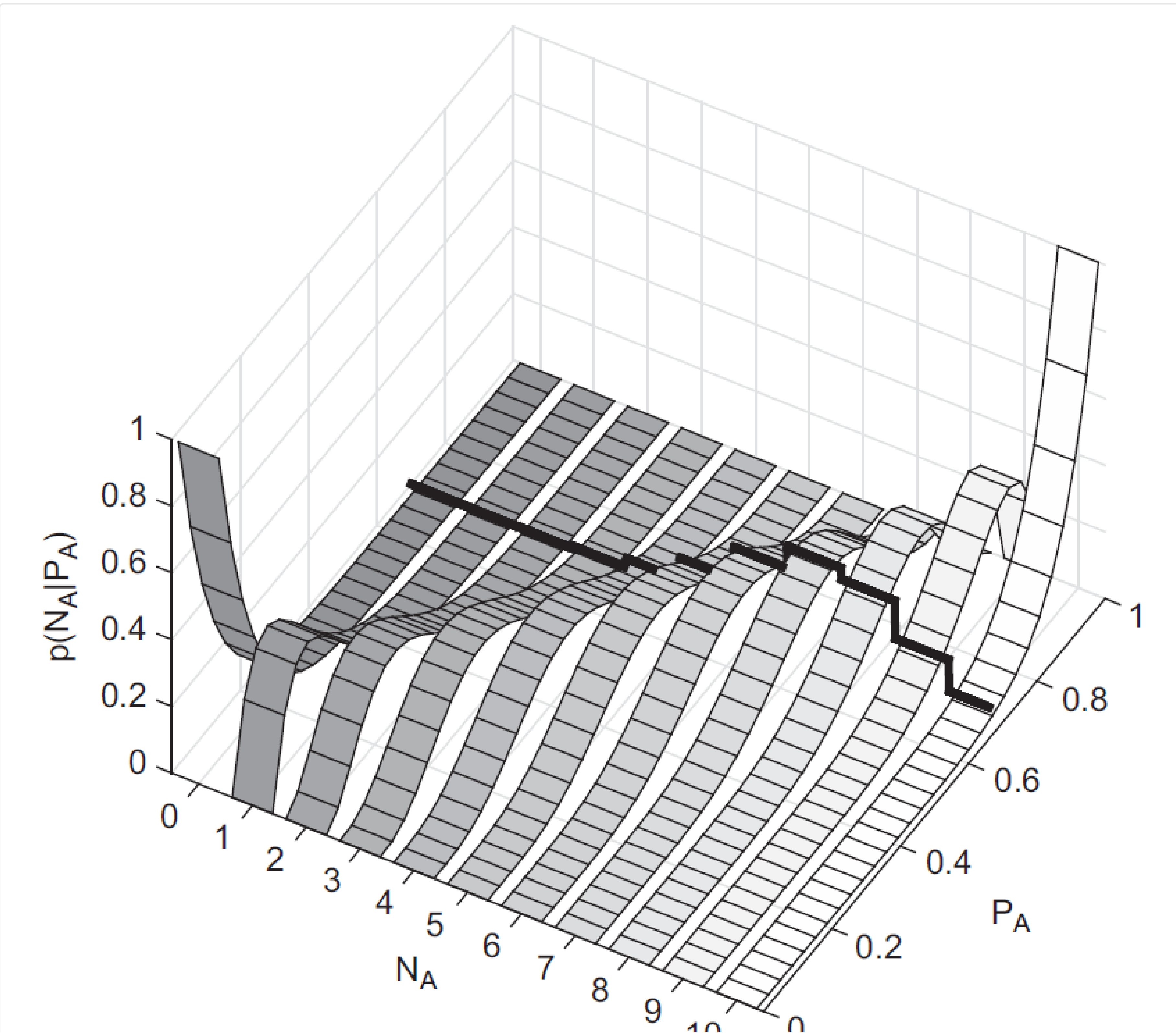


图 7 二项式模型下数据点的概率

定义概率分布

由心理模型所指定的概率函数

漂移 Wald 概率密度函数：

$$f(t \mid a, m, T) = \frac{a}{\sqrt{2\pi(t-T)^3}} \exp\left(-\frac{[a-m(t-T)]^2}{2(t-T)}\right), t > T$$

参数说明：

t： 反应时

m： 漂移（率）

a： 反应边界

T： 启动时间

```R
rswald <- function(t, a, m, Ter){
  ans <- a/sqrt(2*pi*(t-Ter)^3)*
    exp(-(a-m*(t-Ter))^2/(2*(t-Ter)))
}
```

脚本 1 漂移 Wald 概率密度函数

基于数据模型的概率函数

GCM

$$d_{ij} = \left(\sum_{k=1}^{K} \left|x_{ik} - x_{jk}\right|^2\right)^{\frac{1}{2}}$$

$$s_{ij} = \exp\left(-c \cdot d_{ij}\right)$$

$$P\left(R_i = A \mid i\right) = \frac{\left(\sum_{j \in A} s_{ij}\right)}{\left(\sum_{j \in A} s_{ij}\right) + \left(\sum_{j \in B} s_{ij}\right)}$$

```R
source("GCMpred.R")

N <- 2*80
N_A <- round(N*.968)

c <- 4
w <- c(0.19, 0.12, 0.25, 0.45)

stim <- as.matrix(read.table("faceStim.csv", sep=","))

exemplars <- list(a=stim[1:5,], b= stim[6:10,])
```

```r
preds <- GCMpred(stim[1,], exemplars, c, w)

likelihood <- dbinom(N_A ,size = N,prob = preds[1])
```

脚本 2 连接 GCM 模型与二项式函数

```r
GCMpred <- function(probe, exemplars, c, w){

  dist <- list()
  for (ex in exemplars){
    dist[[length(dist)+1]] <- apply(as.array(ex), 1,
                          function(x) sqrt(sum(w*(x-probe)^2)))
  }

  sumsim <- lapply(dist, function(a) sum(exp(-c*a)))

  r_prob <- unlist(sumsim)/sum(unlist(sumsim))

}
```

脚本 3 从 GCM 获取反应的预测概率的代码

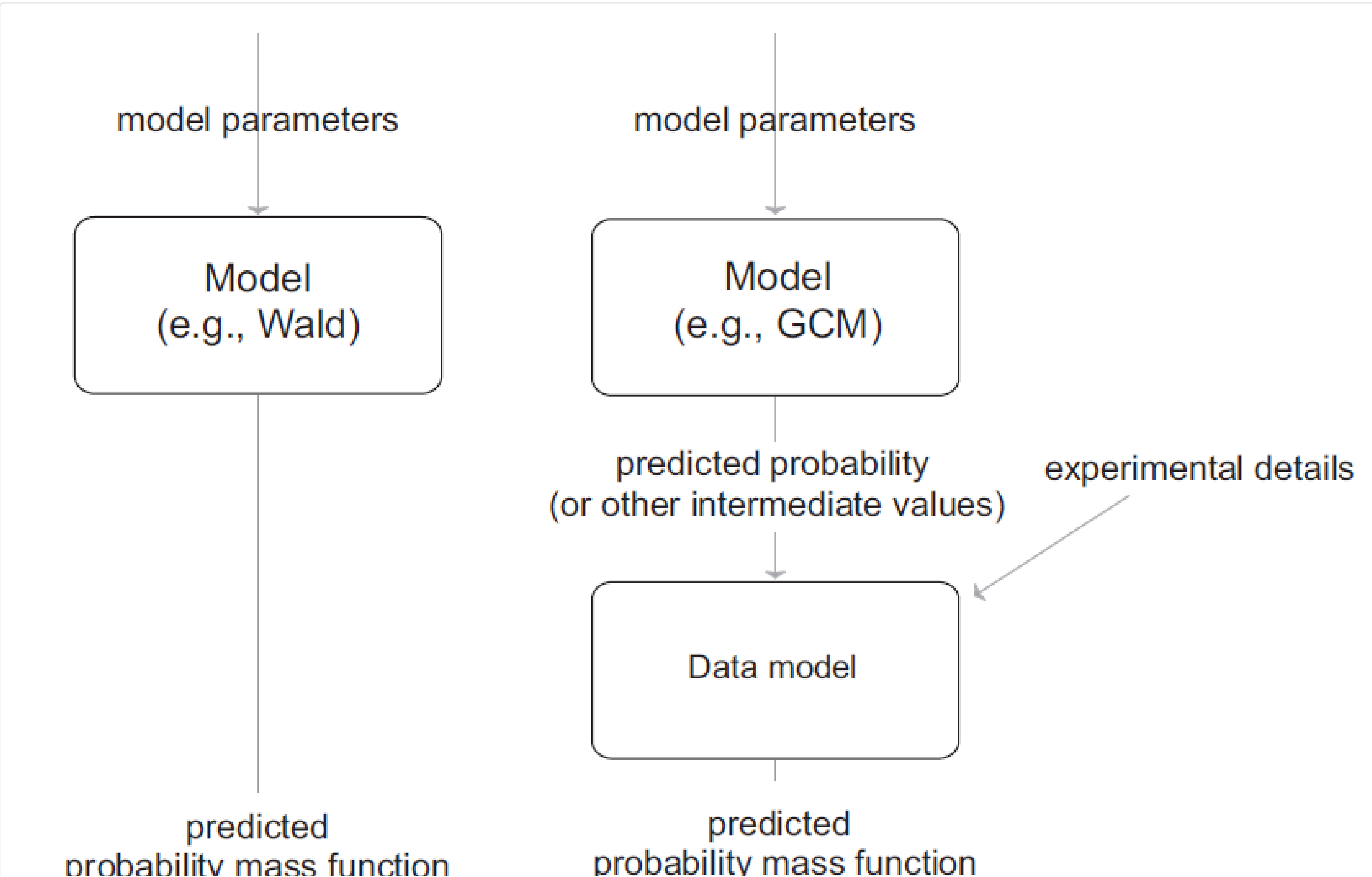
GCM 预测下的二项分布

$$f\left(k \mid p_{\text{heads}}, N\right)=\binom{N}{k} p_{\text{heads}}^{k}\left(1-p_{\text{heads}}\right)^{N-k}$$

$$f\left(N_A \mid P_A, N\right)=\binom{N}{N_A} P_A^{N_A}\left(1-P_A\right)^{N-N_A}$$

概率函数的两种类型

1. 模型参数和模型一起足以预测全概率函数
2. 模型参数和模型一起预测某些中间变量，再与其他实验细节一起，利用数据模型指定全概率函数



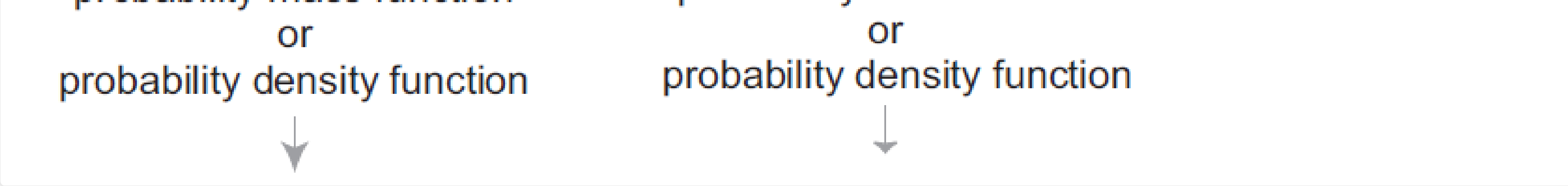


图 8 生成预测概率函数的两种不同方法

扩展数据

二项式分布→多项式分布：

$$
f(\mathbf{N} \mid \mathbf{p}, N_T) = \frac{N_T!}{N_1! N_2! \ldots N_J!} p_1^{N_1} p_2^{N_2} \ldots p_J^{N_J}
$$

多个观测点的似然：

$$
L(\boldsymbol{\theta} \mid \mathbf{y}) = \prod^{k} L(\boldsymbol{\theta} \mid y_k)
$$

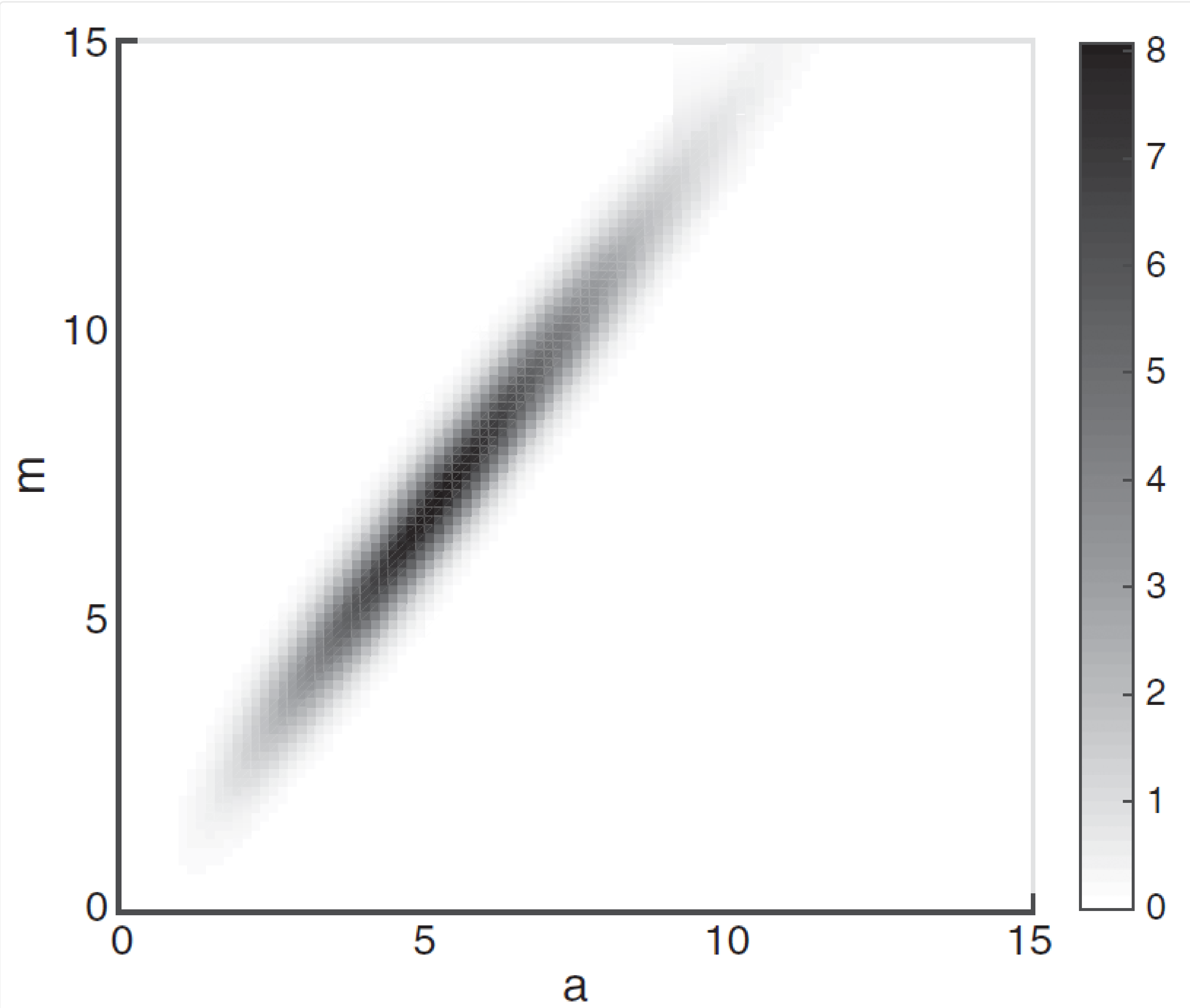


图 9 给定数据集 t = [0.6, 0.7, 0.9]的 Wald 参数 m 和 a 的联合似然函数

寻找最大似然

一般而言，取自然对数方便计算

$$\ln \prod_{k=1}^{K} f(k) = \sum_{k=1}^{K} \ln f(k)$$

$$\ln L(\boldsymbol{\theta} \mid \mathbf{y}) = \sum_{k=1}^{K} \ln L\left(\boldsymbol{\theta} \mid y_k\right)$$

以正态分布作为例子：

$$p(y \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

$$\ln L(\mu, \sigma \mid y) = \ln(1) - \ln\left(\sqrt{2\pi\sigma^2}\right) - \frac{(y-\mu)^2}{2\sigma^2}$$

若我们只关心 μ 的取值，在似然函数中只需要关心最后一项。

后续将使用 $-2lnL$ 来评估模型的拟合度

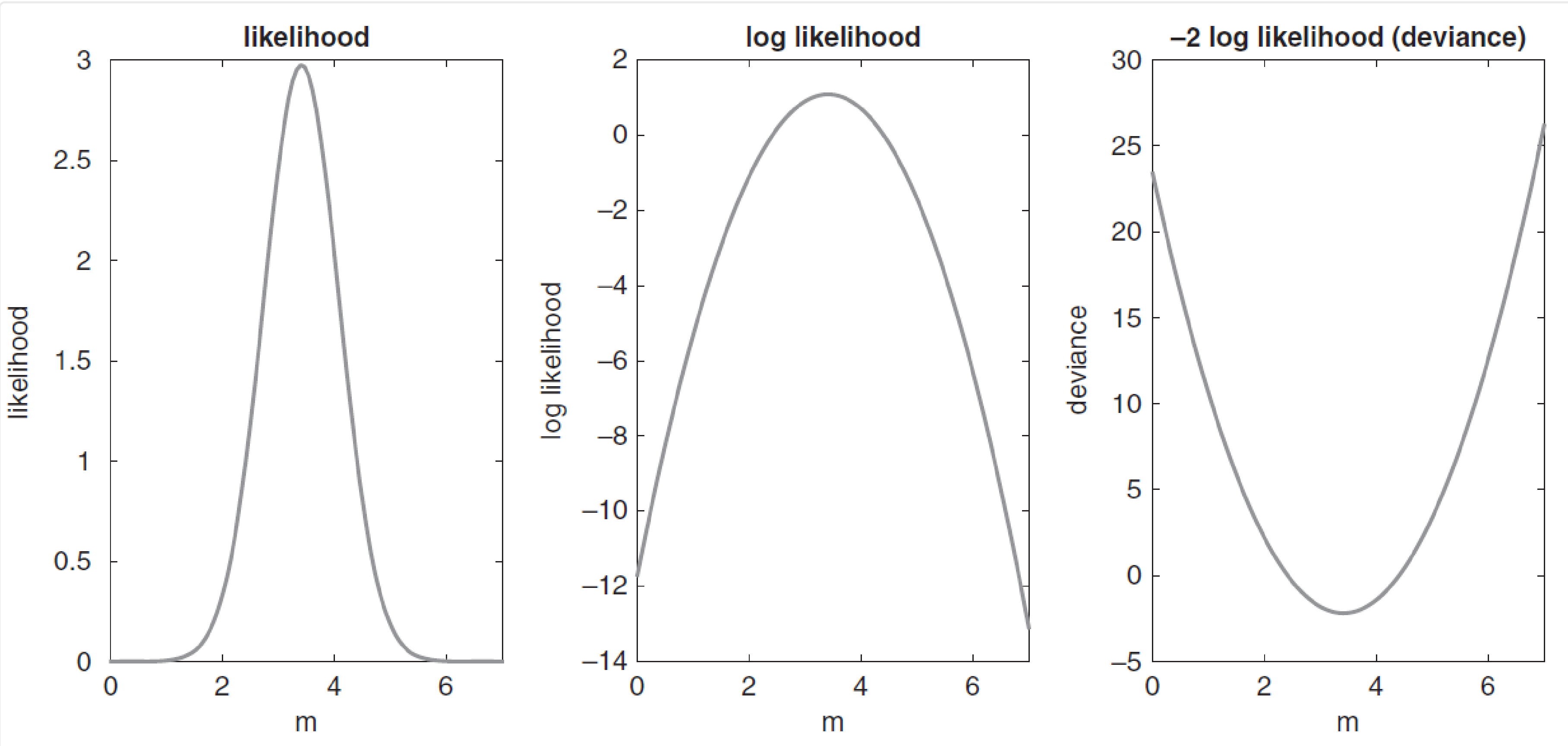


图 10 似然函数、对数似然函数、偏差函数

```
GCMprednoisy ← function(probe, exemplars, c, w, sigma, b){

  dist ← list()
  for (ex in exemplars){
    dist[[length(dist)+1]] ← apply(as.array(ex), 1,
                                   function(x) sqrt(sum(w*(x-probe)^2)))
  }

  sumsim ← unlist(lapply(dist, function(a) sum(exp(-c*a))))

  r_prob ← c(0,0)
  r_prob[1] ← pnorm(sumsim[1]-sumsim[2]-b,sd=sigma)
  r_prob[2] ← 1 - r_prob[1]
```

```r
  return(r_prob)

}
```

脚本 4 用确定性反应规则实现 GCM 版本的 R 代码

```r
source("GCMprednoisy.R")
library(dfoptim)

# A function to get deviance from GCM
GCMutil <- function(theta, stim, exemplars, data, N, retpreds){
  nDat <- length(data)
  dev <- rep(NA, nDat)
  preds <- dev

  c <- theta[1]
  w <- theta[2]
  w[2] <- (1-w[1])*theta[3]
  w[3] <- (1-sum(w[1:2]))*theta[4]
  w[4] <- (1-sum(w[1:3]))
  sigma <- theta[5]
  b <- theta[6]

  for (i in 1:nDat){
    p <- GCMprednoisy(stim[i,], exemplars, c, w, sigma, b)
    dev[i] <- -2*log(dbinom(data[i] ,size = N,prob = p[1]))
    preds[i] <- p[1]
  }

  if (retpreds){
    return(preds)
  } else {
    return(sum(dev))
  }
}

N <- 2*40

stim <- as.matrix(read.table("faceStim.csv", sep=","))

exemplars <- list(a=stim[1:5,], b= stim[6:10,])

data <- scan(file="facesDataLearners.txt")
data <- ceiling(data*N)

bestfit <- 10000

for (w1 in c(0.25,0.5,0.75)){
  for (w2 in c(0.25,0.5,0.75)){
    for (w3 in c(0.25,0.5,0.75)){
      print(c(w1,w2,w3))
      fitres <- nmkb(par=c(1,w1,w2,w3,1,0.2),
           fn = function(theta) GCMutil(theta,stim,exemplars,data, N, FALSE),
           lower=c(0,0,0,0,0,-5),
           upper=c(10,1,1,1,10,5),
           control=list(trace=0))
      print(fitres)
      if (fitres$value<bestfit){
        bestres <- fitres
        bestfit <- fitres$value
      }
    }
  }
```

```r
}

preds <- GCMutil(bestres$par,stim,exemplars,data, N, TRUE)

plot(preds,data/N,
     xlab="Data", ylab="Predictions")

print(bestres)
theta <- bestres$par
w <- theta[2]
w[2] <- (1-w[1])*theta[3]
w[3] <- (1-sum(w[1:2]))*theta[4]
w[4] <- (1-sum(w[1:3]))
print(w)
```

脚本 5 对一些数据进行拟合的 GCM 修改版本 R 代码

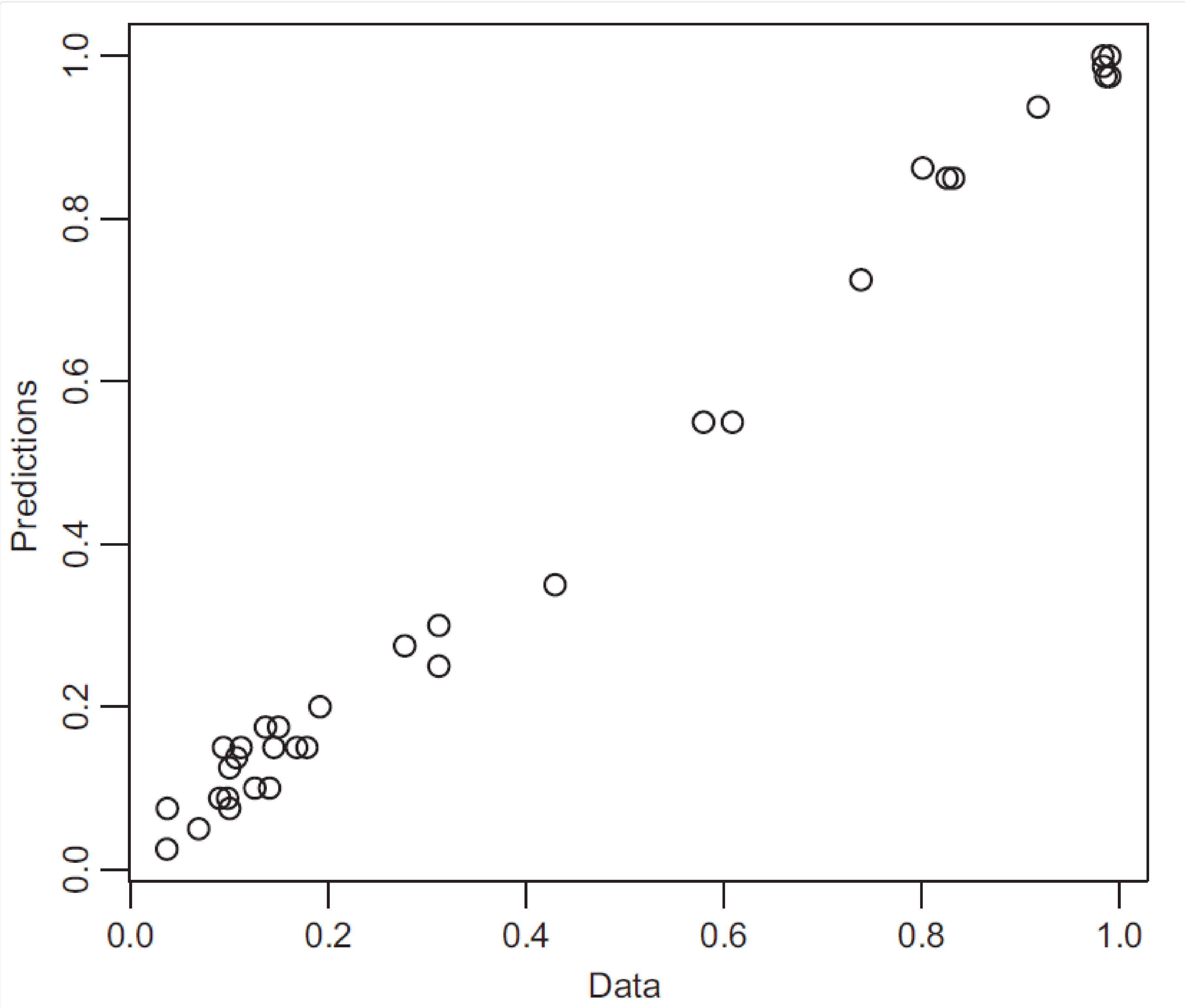


图 11 单个数据与最大似然参数估计下的 GCM 预测概率之间的散点图

最大似然估计量的性质

1. 充分性。意味着，给定统计模型和需要估计的参数 θ， θ 的最大似然估计包含样本提供的有关 θ 的所有信息
2. 参数不变性。即，若存在变换函数 g，则求 $g(\theta)$ 的最大似然，等同于找到 θ 的最大似然，然后应用变换 g
3. 一致性和效率。一致性意味样本越多，估计值越接近真实值；效率意味，随着样本增加，最大似然估计接近于正态分布
4. 通常不具备无偏性
5. 在估计值过于分散的情况下，采用分层模型结构限制部分参数变异性。