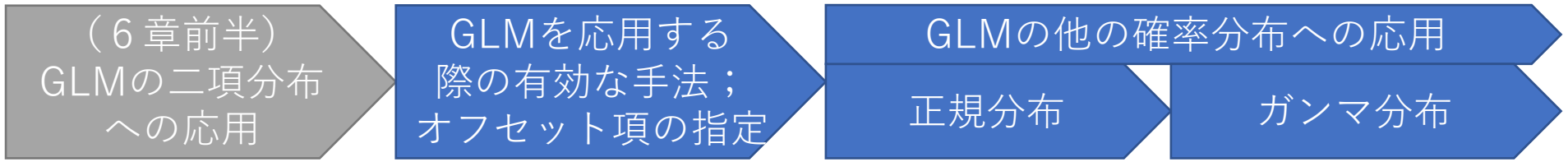
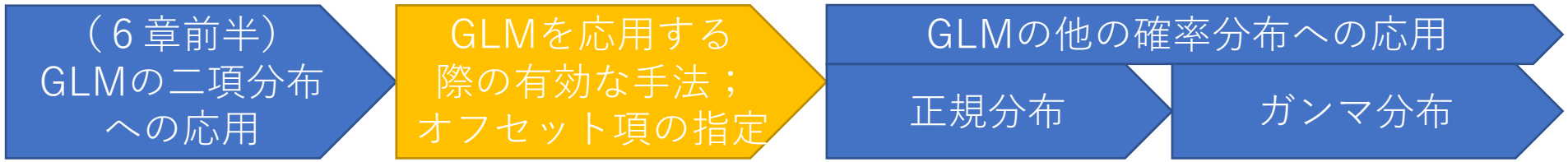
6章：GLMの応用範囲をひろげる

後半　二項分布以外への応用

太田研究室４年　和田



割算値の統計モデリングはやめよう



割算値の統計モデリングは、必ずしも現実を反映しない

（そもそもなぜ、二項分布やロジスティック回帰を使うのか？）

観測値の誤った「解析」が起こりがち (以下例)

- 生起確率＝（観測データ）/（観測データ）
- 観測値の変数変換　　e.g. 対数変換, 平均化, etc.

問題

- 1000回中300回起こる現象と、10回中3回起こる現象は、同じ確からしさか？
- 分子の観測データ、分母の観測データにそれぞれ誤差があるとき…生起確率はどんな確率分布になる？

N_1個中y_1個の種子が生存するとき…

「生存確率を y_1/N_1 の単純計算で求める」

$\neq$

「生存確率p_1を、$p=p_1$, $N=N_1$, $x=y_1$の二項分布に従うと仮定して探る」

オフセット項を使えば、単純比例の関係もGLMに組み込める

- ~~観測値を割り算や変数変換→得た数値を応答変数に~~　間違い
- 「人口密度」などを調べたい＝観測データと観測データの比例関係が知りたいときはどうすれば…？　→　オフセット項設置

架空データ・調査
・森林内に調査値を100箇所設置
・調査値iごとに面積A_iが異なる
・調査値iの「明るさ」x_iを測定
・調査値iにおける植物個体数y_iを記録
・明るさx_iは植生密度にどう影響しているかを調査

$$\frac{\text{平均個体数}\lambda_i}{A_i} = \text{人口密度}$$

モデル化

$$\lambda_i = A_i \times \text{人口密度} = A_i \exp(\beta_1 + \beta_2 x_1)$$

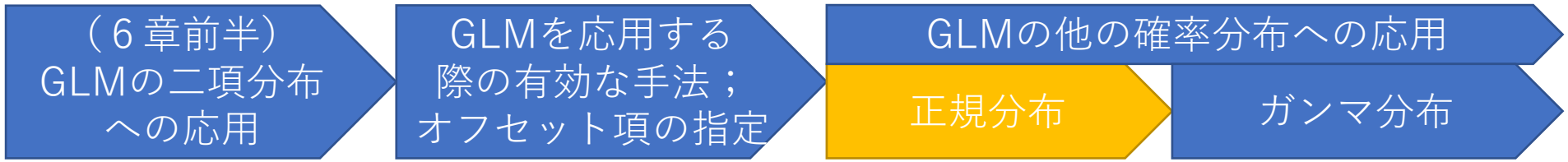
$$\lambda_i = \exp(\beta_1 + \beta_2 x_1 + logA_i)$$

線形予測子：　$z_i = \beta_1 + \beta_2 x_1 + logA_i$

```
> glm(y~x, offset=log(A), family=poisson, data=d)
```

・個体数平均は調査値の面積に比例
・明るさx_iの影響を推定可能

正規分布とその尤度

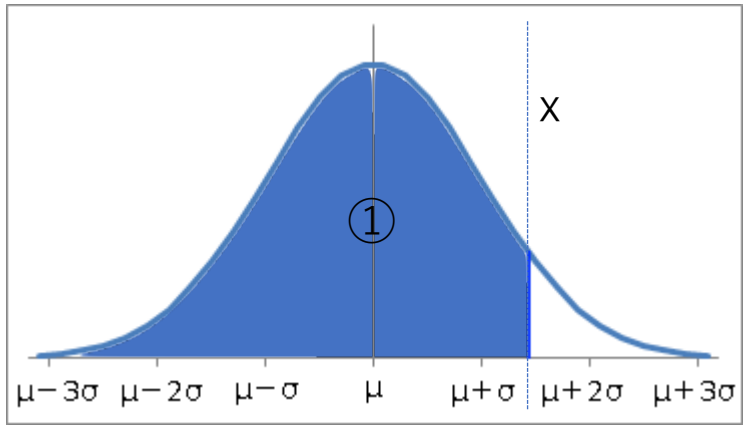

（6章前半）
GLMの二項分布
への応用
GLMを応用する
際の有効な手法；
オフセット項の指定
GLMの他の確率分布への応用
正規分布
ガンマ分布

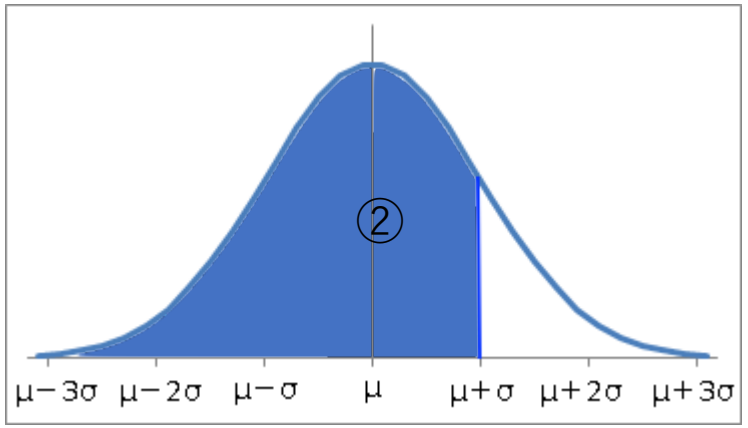
正規分布を扱う際にも最尤推定を行うことができる①

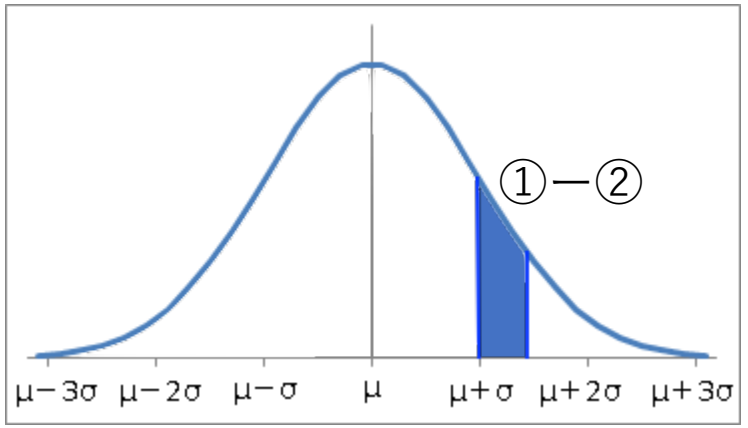
正規分布のおさらい
・連続値のデータを扱う
・範囲は-∞ ～ +∞
・パラメーターは二つ
　平均値 μ
　標準偏差 σ

確率密度関数

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

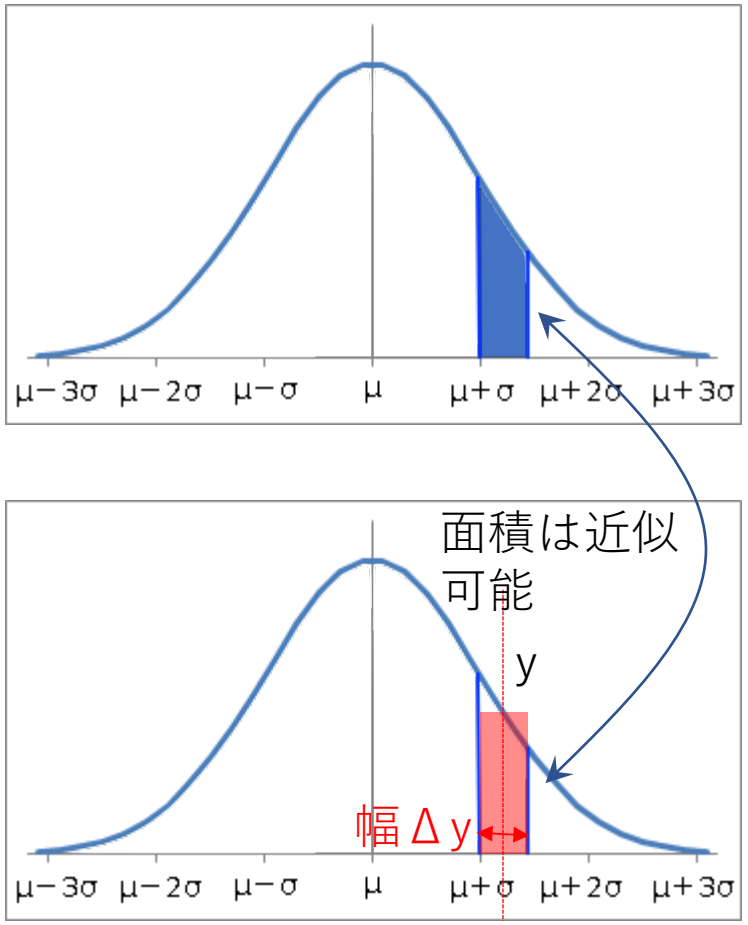






pnorm(x, μ , σ)で求められるのは青い面積

正規分布を扱う際にも最尤推定を行うことができる②



尤度関数は

$$L(\mu,\sigma)=\prod_i p(yi|\mu,\sigma)\Delta y=\prod_i \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}\Delta y$$

$$logL(\mu,\sigma)=-0.5Nlog(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_i(y_i-\mu)^2+Nlog(\Delta y)$$

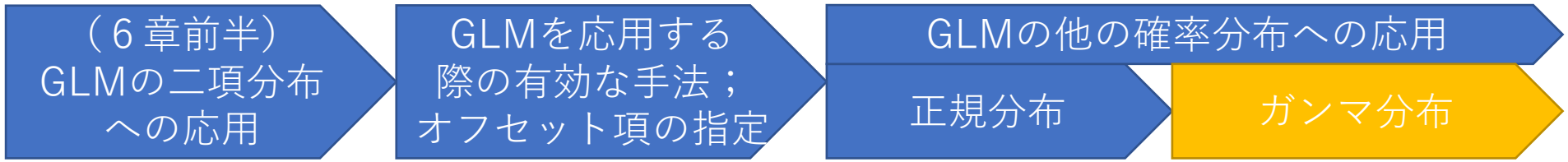
σはμと無関係の定数とする

Δyは定数

このσの項目が最小なるとき
（$=(y_i-\mu)^2$が最小のとき）
logL(μ, σ)が最大

「最小二乗法」

ガンマ分布のGLM

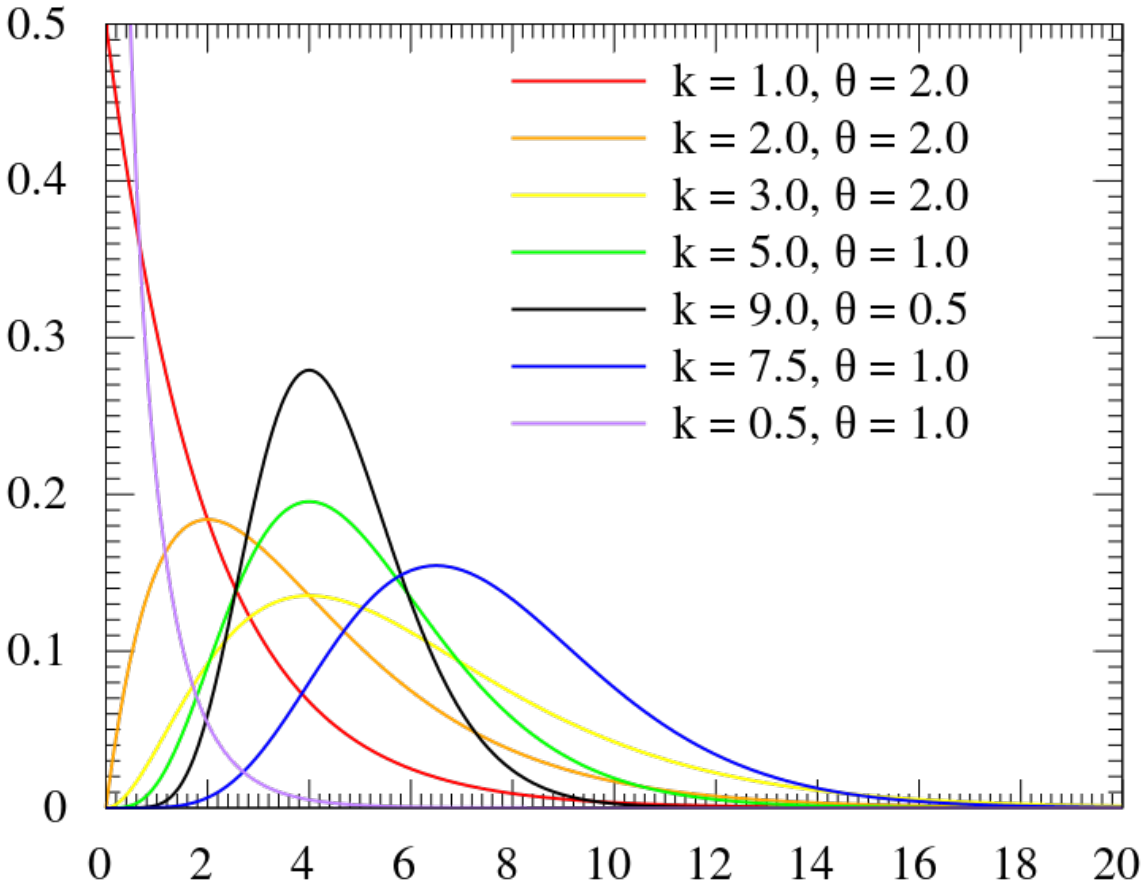

（6章前半）
GLMの二項分布
への応用
GLMを応用する
際の有効な手法；
オフセット項の指定
GLMの他の確率分布への応用
正規分布
ガンマ分布

ガンマ分布とは何か

- 期間 θ に１回起こる現象がk回起こるまでにかかる時間の分布

右図　オレンジの線

- 2分に一度電話が鳴るコールセンター
- 出勤してから、電話が2度鳴るまでにかかる時間
- 1分で2度鳴る確率が約15％
- 2分で2度鳴る確率が約19％
- 4分で2度鳴る確率が8％



ガンマ分布の特徴

確率密度関数：$f(x) = x^{k-1}\frac{\exp(-\frac{x}{\theta})}{(k-1)!\,\theta^k}$

非負の連続確率分布

パラメーターは２つ（k, θ ）

平均はk θ, 分散はk θ^2