

無線システムの電波伝搬理解に 役立つ確率・統計モデル

唐沢 好男

自己紹介; 信学会通ソマガジン B-plus

https://www.jstage.jst.go.jp/article/bplus/16/3/16_238/_pdf

内容

第一部：確率分布

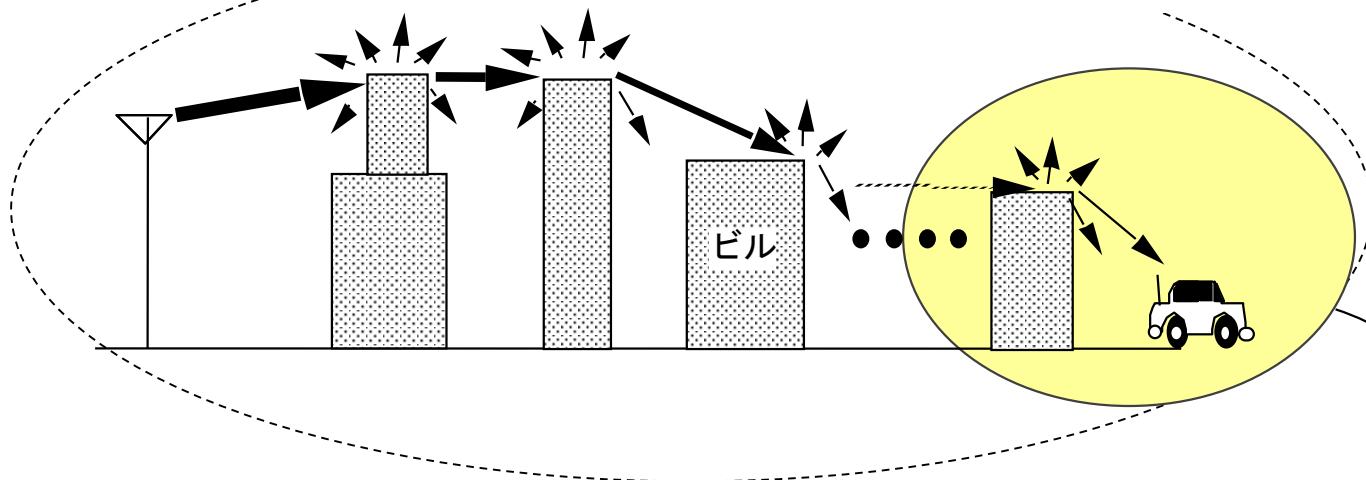
- ・無線通信と電波伝搬
- ・確率過程
 - 加法性確率過程と正規分布
 - 乗法性確率過程と対数正規分布
- ・移動通信の電波伝搬モデル
 - レイリー分布
 - 仲上・ライス分布
 - 仲上 m 分布

第二部：統計的推定

- ・地球温暖化時代において雨の降り方は変わってきたのか？
を一緒に考えてゆこう(降雨は無線通信の重要問題)
- ・統計的推定手法
 - 回帰直線と区間推定：誤った結論を出さないために
 - 回帰曲線と赤池情報量基準(AIC)：より良いモデルを見つけるために
 - 極値統計：想定外の出来事を想定外にしないために

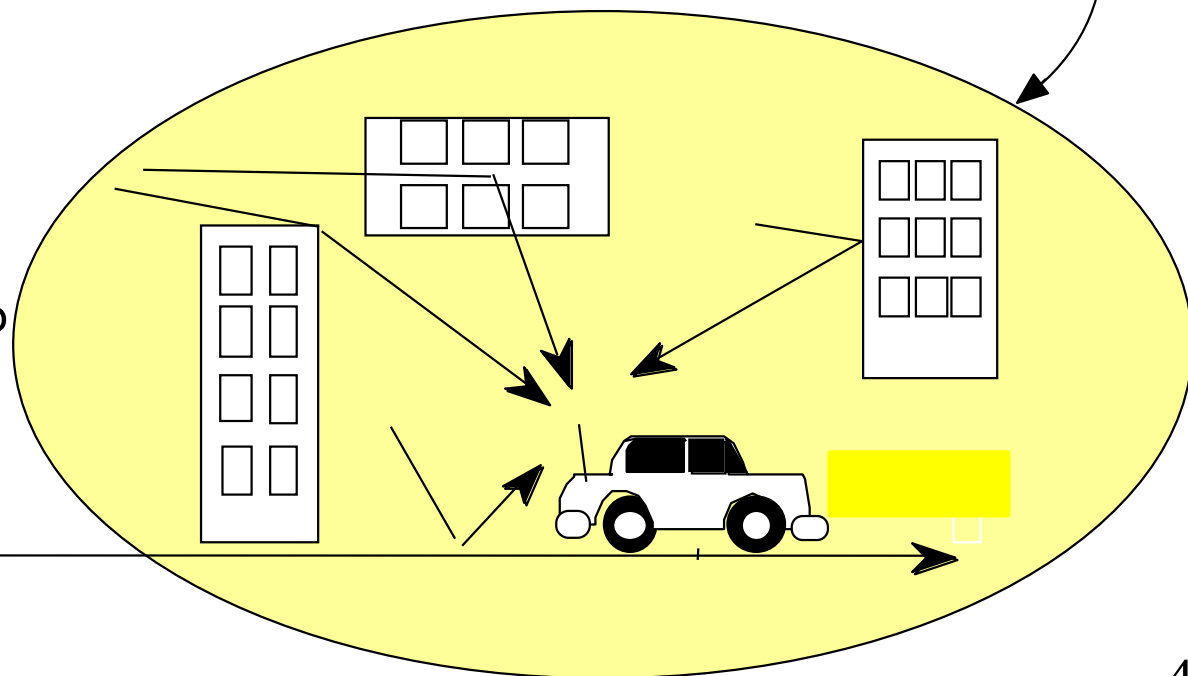
移動通信の電波伝搬

電波が弱くなる

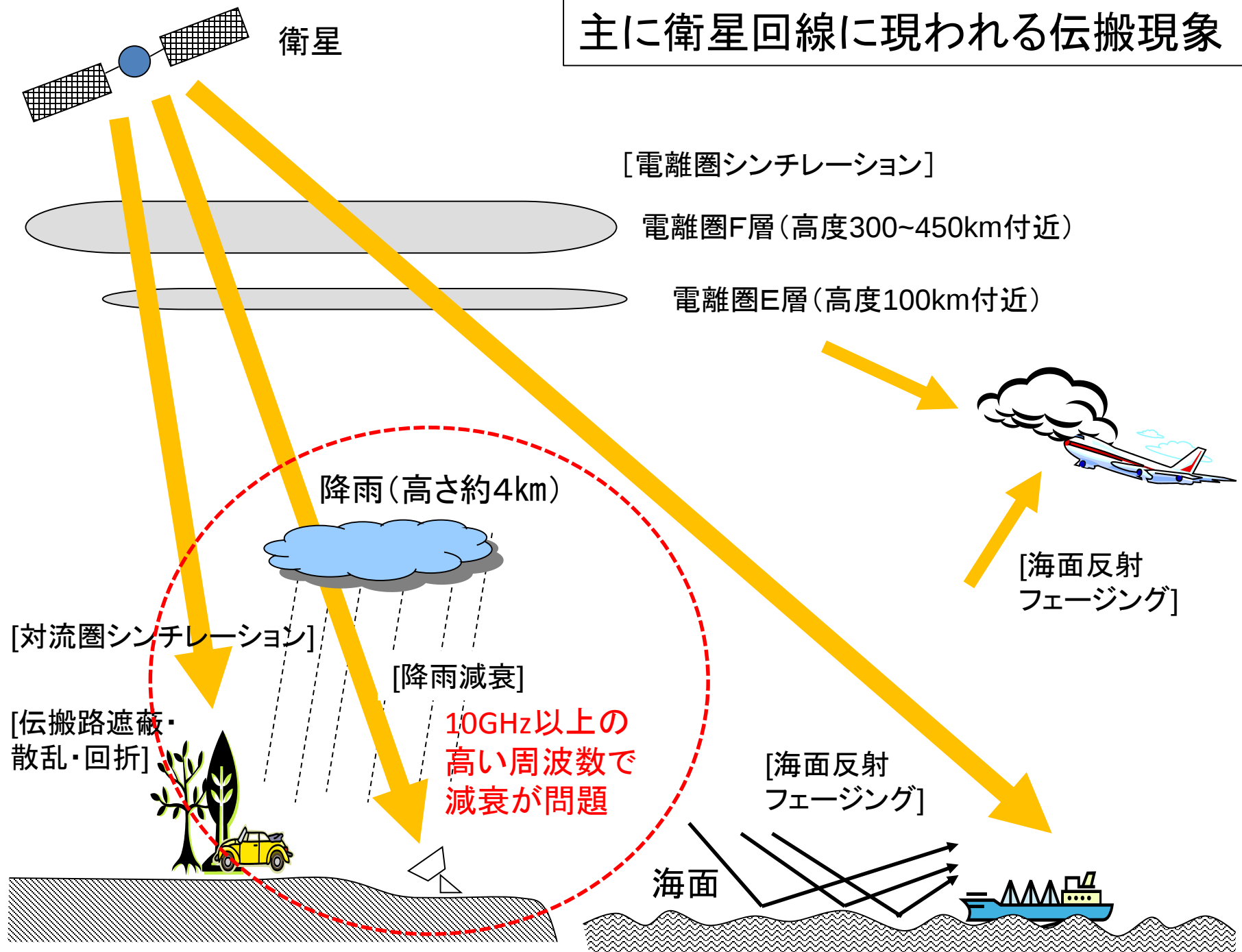


マルチパスにより
様々な信号劣化が起きる

- ・電波の強度低下
- ・速い変動
- ・伝送波形の歪み



主に衛星回線に現われる伝搬現象



確率と確率分布(確率密度関数と累積分布関数)

確率 ある事象の起こりやすさ(あるいは、確からしさ)の指標: 値は 0 ~ 1

分かりやすい例: 事象が限られた数

- ・コインの裏表(1/2)、さいころの目(1/6)

分かりにくい例: 事象が無限にある(=連続量)

- ・明日の最高気温が15°Cになる確率 (答えられない)
- ・明日の最高気温が15°C以上になる確率
(条件が与えられれば答えられる)

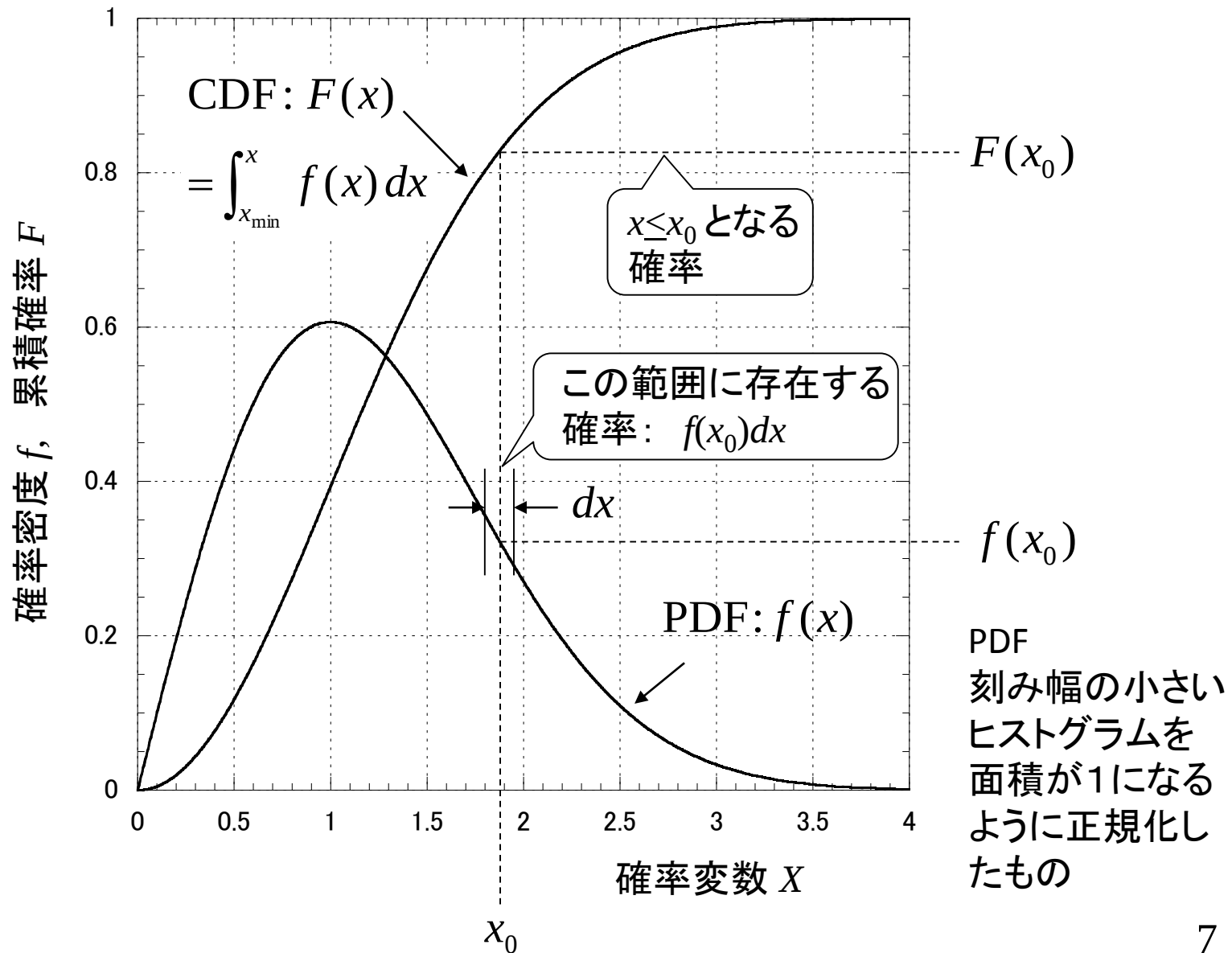
確率密度関数 (PDF) $f(x)$

確率変数 X が、微小区間 $x - \frac{\Delta x}{2} \leq X \leq x + \frac{\Delta x}{2}$ にある確率: $f(x)\Delta x$

累積分布関数 (CDF) $F(x)$

確率変数 X が $X \leq x$ である確率 $F(x) = \int_{-\infty}^x f(x) dx$

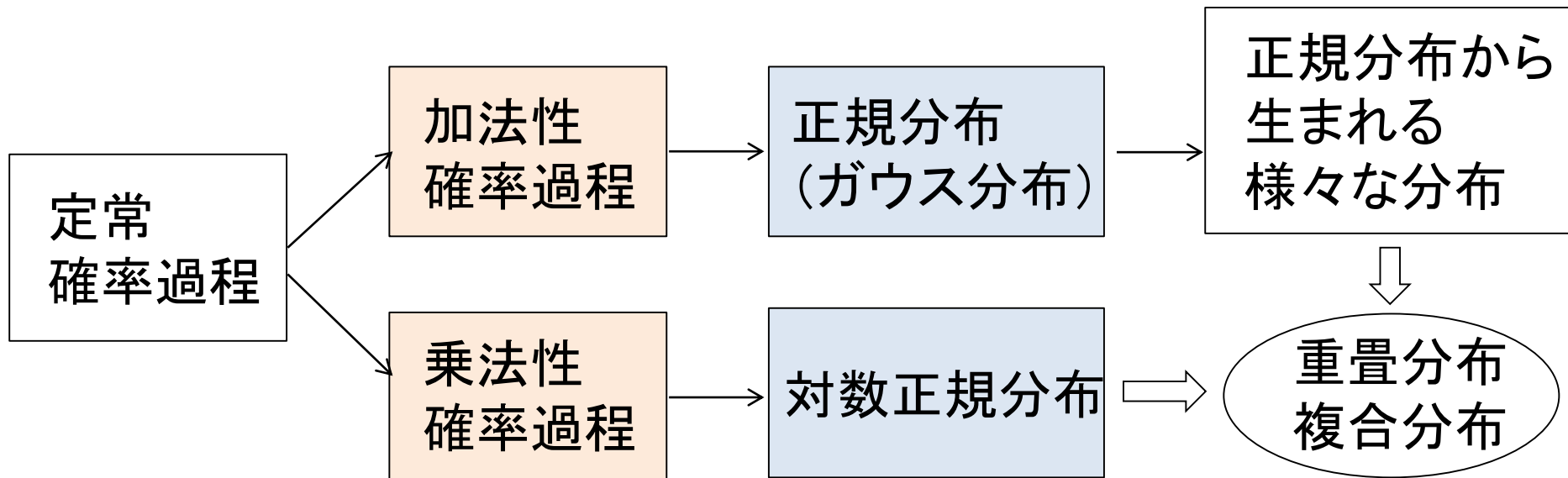
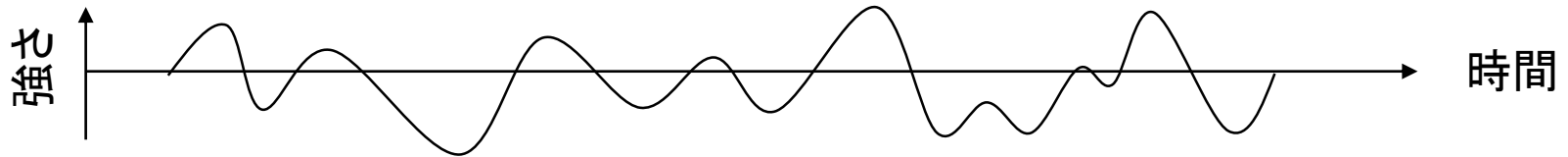
確率密度関数(PDF)と累積分布関数(CDF)



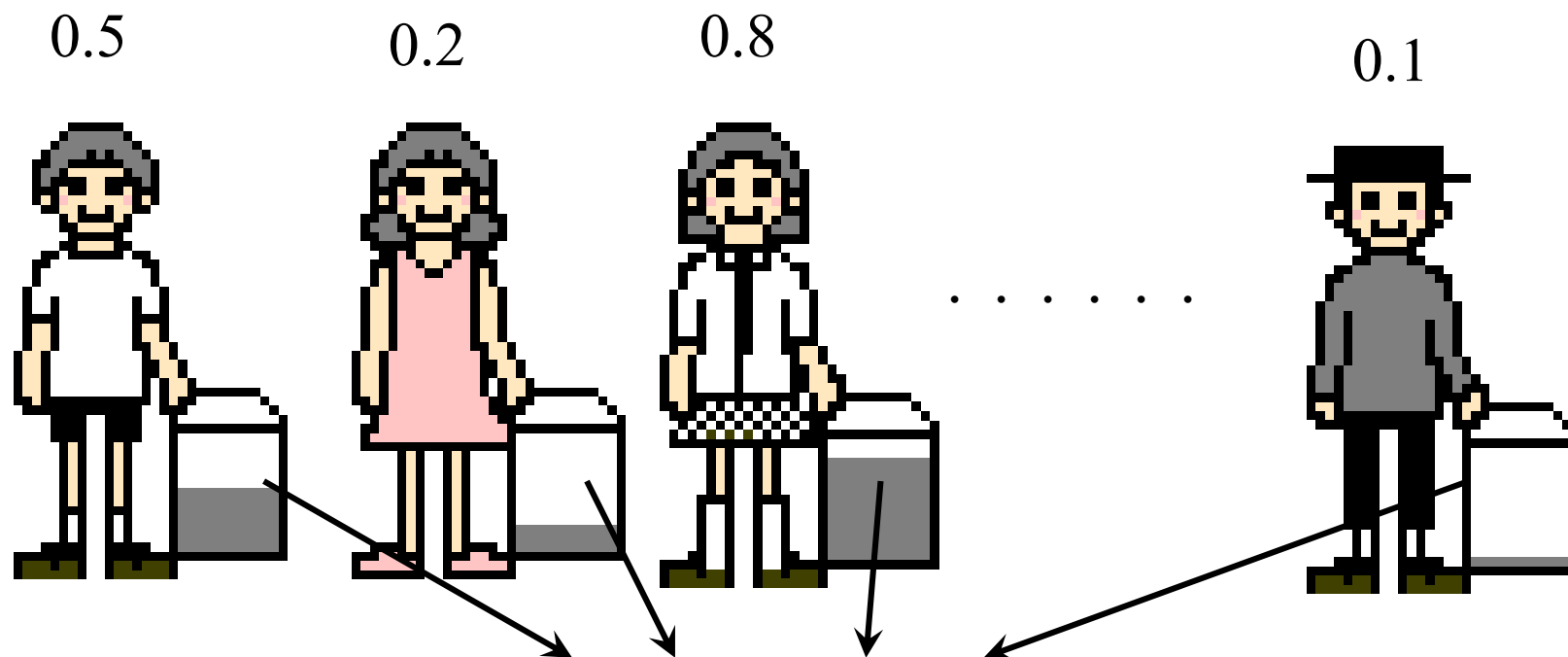
移動伝搬と確率過程

電波伝搬： 電波と自然現象との関わりを扱う分野

不規則媒質中の電波伝搬 → 確率過程



正規分布(ガウス分布、Normal distribution, Gaussian distribution) : 加法性確率過程 (Additive Stochastic Process)



中心極限定理
Central limit theorem

正規分布
 $N(m, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-m)^2}{2\sigma^2} \right\}$$

中心極限定理 (Central Limit Theorem)

$$\hat{X} = (X_1 + X_2 + \cdots + X_N) / N$$



x_i : 同一かつ独立な分布 (i.i.d.)

(もう少し条件を緩めることも可能)

N が十分大きいとき、正規分布になる

個々の作用が足し算される現象の物理量は正規分布に
漸近する $\Rightarrow$ **正規分布は確率分布の基本中の基本分布**

この証明は、どの確率の教科書にも書かれていて、さほど難しくない。
しかし、その証明を理解するためには、確率分布の和の分布を求める
特性関数(あるいは積率母関数)の知識が必要

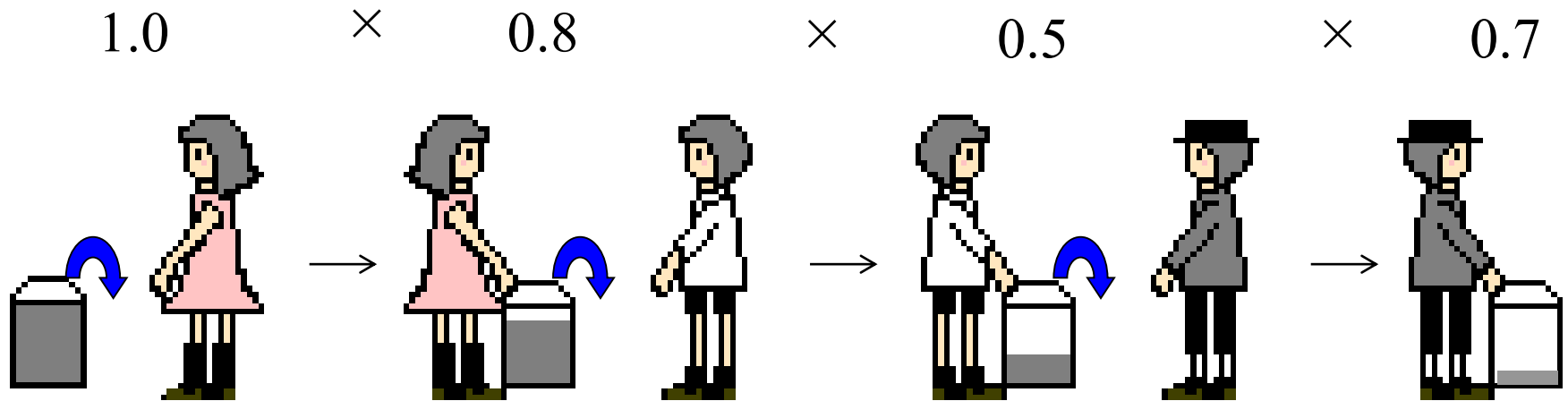
対数正規分布 (log-normal distribution)

乗法性確率過程 (Multiplicative Stochastic Process)

$$x = x_1 \cdot x_2 \cdot x_3 \cdots x_N$$

$$\log x = \log x_1 + \log x_2 + \log x_3 + \cdots + \log x_N$$

正規分布



対数正規分布

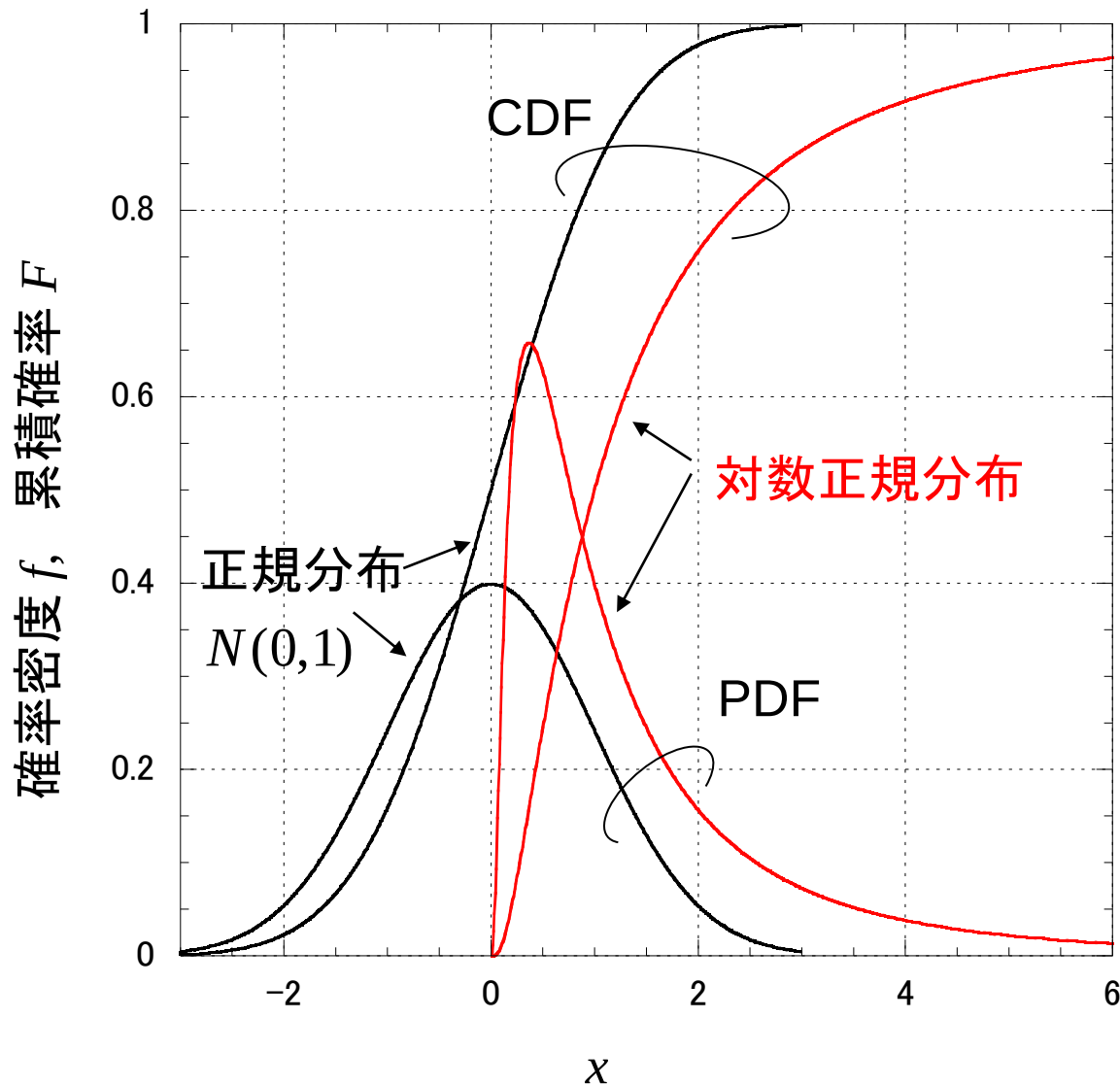
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp \left\{ -\frac{(\log x - m)^2}{2\sigma^2} \right\}$$

対数で表された量 (例えばdB値)
が正規分布するときの

真数値 x の分布 (誤差を倍半分で見える世界)

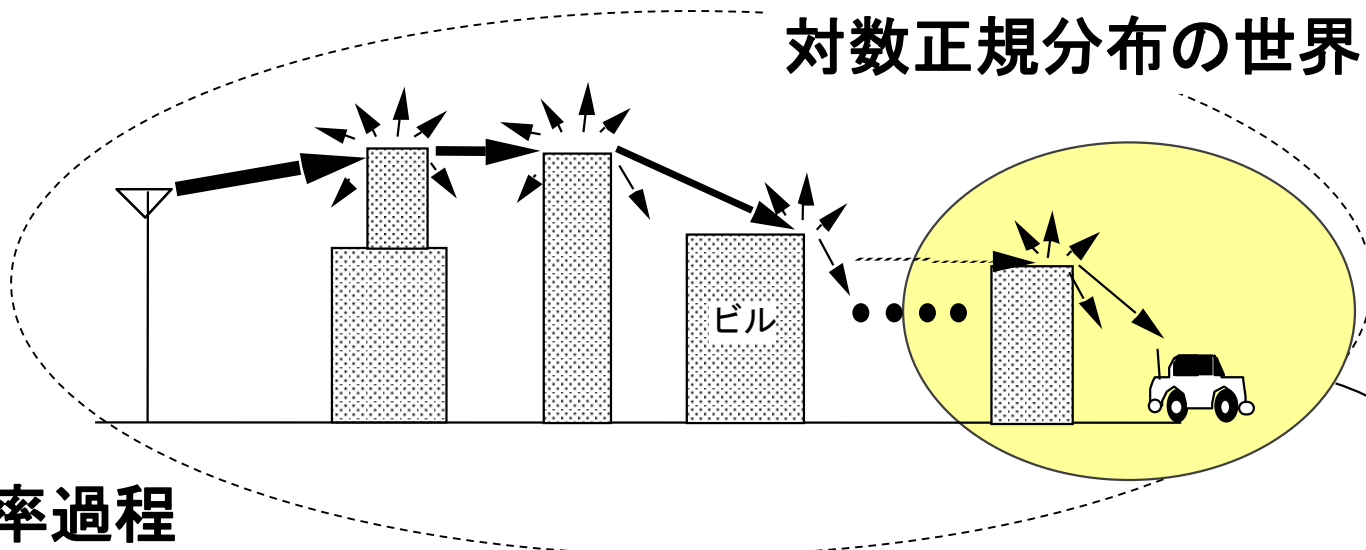
(m, σ は $\log x$ の平均値と標準偏差) 11

正規分布と対数正規分布の 確率密度関数(PDF)と累積分布関数(CDF)



移動伝搬の確率過程

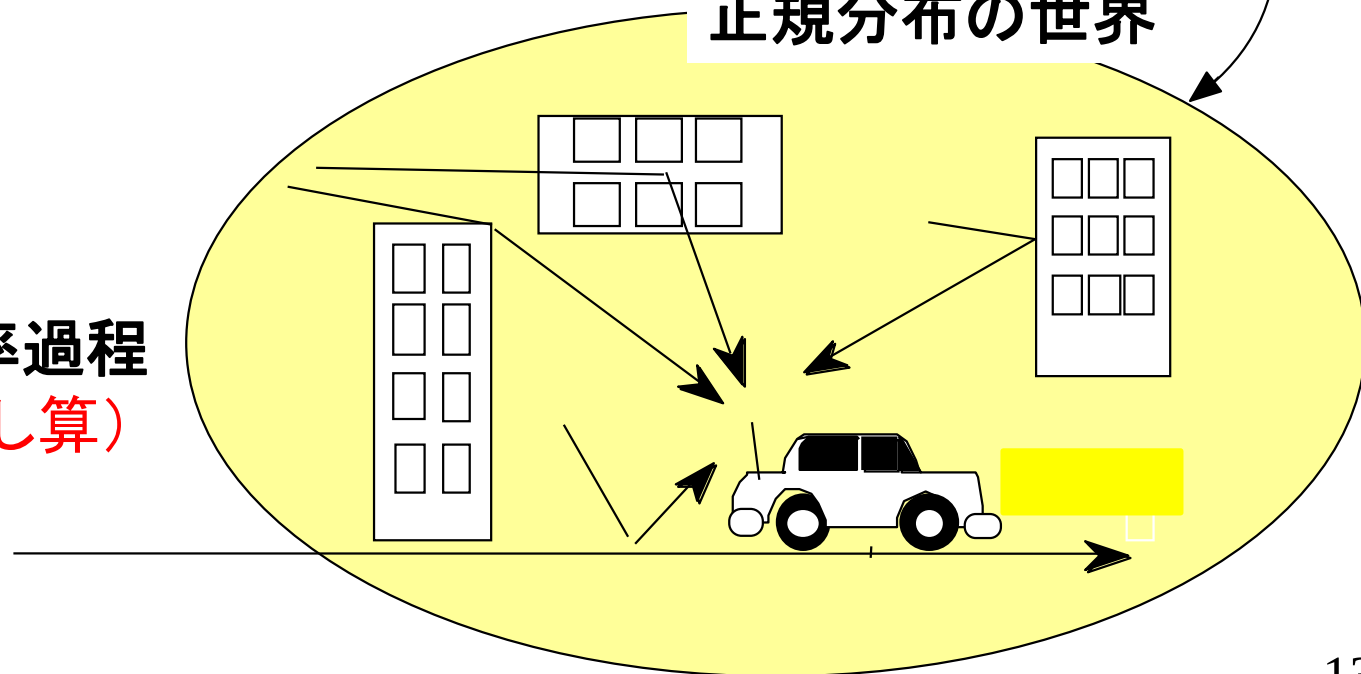
対数正規分布の世界



乗法性確率過程
(現象が掛け算)

正規分布の世界

加法性確率過程
(現象が足し算)



移動伝搬モデルに現れる基本確率分布

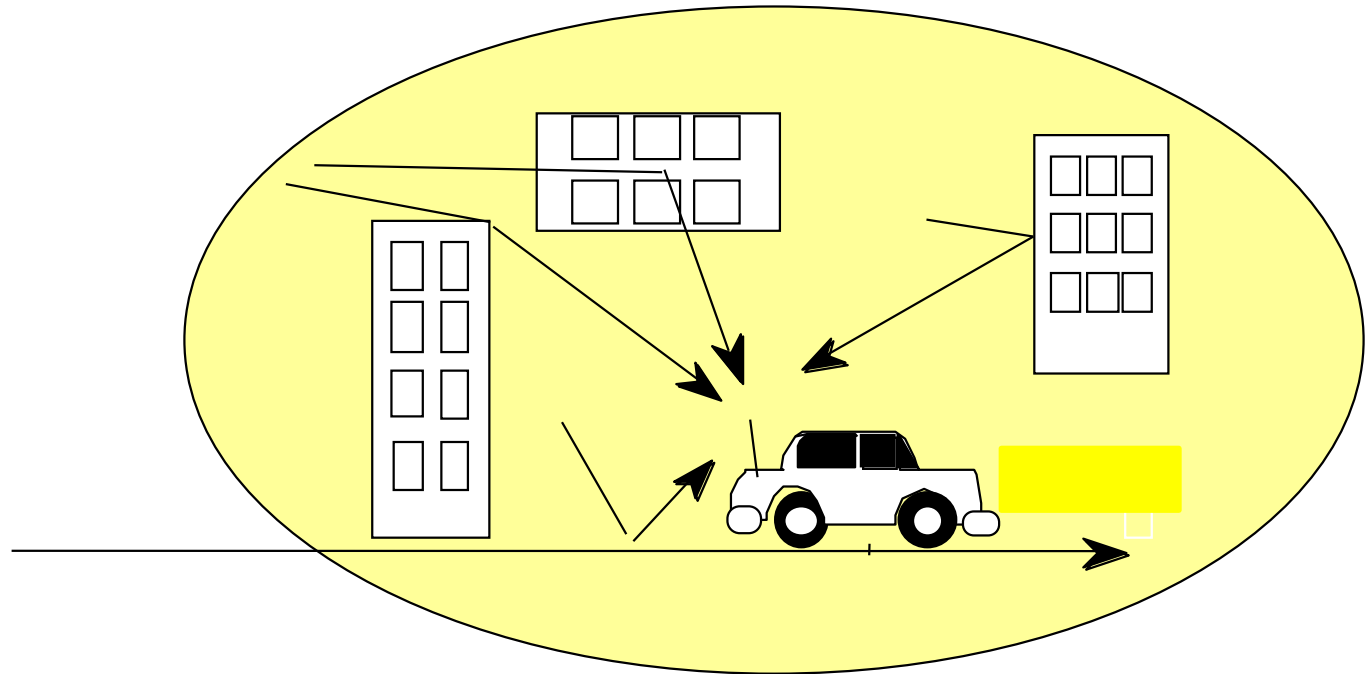
通信モデルと確率分布

物理量	代表的確率分布
雑音 信号強度	正規分布 レイリー分布、(NLOS) 仲上・ライス分布、(LOS) Loo分布、(SLOS) 仲上m分布
待ち時間	指数分布
発生回数	ポアソン分布
ダイバーシティ	ガンマ分布(カイニ乗分布)
MIMO	ウィシャート分布
遮へい減衰	対数正規分布
・	・
・	・

(各種確率分布の文献情報:最後のスライドに)

マルチパスフェージングの振幅変動を表す確率分布

加法性確率過程



- ① レイリー分布
- ② 仲上・ライス分布
- ③ 仲上 m 分布

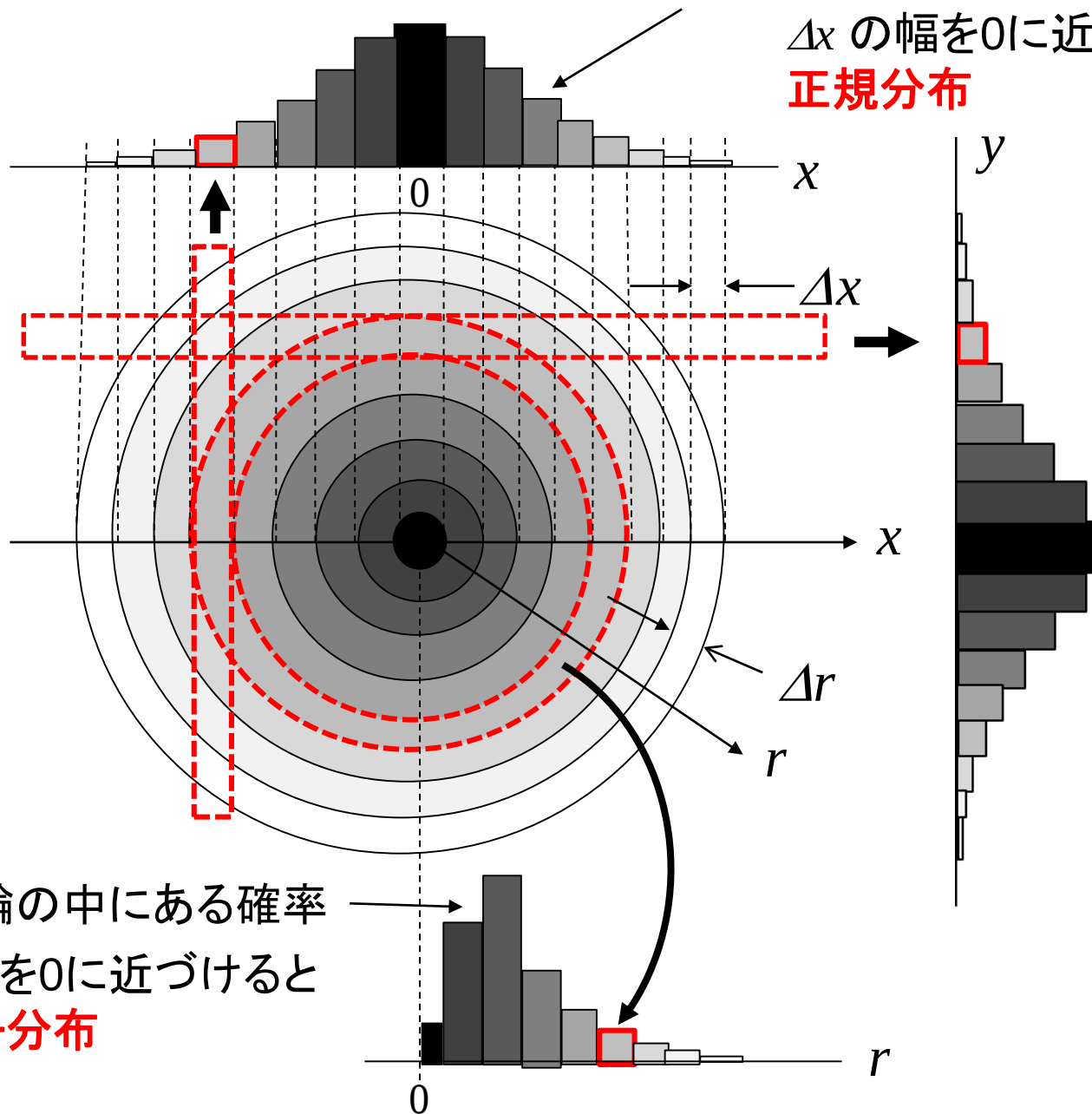
① レイリー分布

見通し外のフェージング環境(レイリーフェージング環境)を表す

移動通信の基本分布

$r=0$ で確率密度が0である意味

Δx の幅の矩形の中にある確率
 Δx の幅を0に近づけると
正規分布



Δr の幅の輪の中にある確率
 Δr の幅を0に近づけると
レイリー分布

(Lord Rayleigh, 1880)

電波環境では： 同レベルの多数の波が合成された信号の包絡線レベルの分布
(見通し外(NLOS)伝搬環境)

$$\begin{aligned}
 a &= r_1 e^{j\phi_1} + r_2 e^{j\phi_2} + \cdots + r_N e^{j\phi_N} \\
 &= \underbrace{r_1 \cos \phi_1 + r_2 \cos \phi_2 + \cdots + r_N \cos \phi_N}_{x = r \cos \phi \rightarrow N(0, \sigma^2)} + j \underbrace{(r_1 \sin \phi_1 + r_2 \sin \phi_2 + \cdots + r_N \sin \phi_N)}_{y = r \sin \phi \rightarrow N(0, \sigma^2)}
 \end{aligned}$$

$$r = |x + jy| \left(= \sqrt{x^2 + y^2} \right)$$

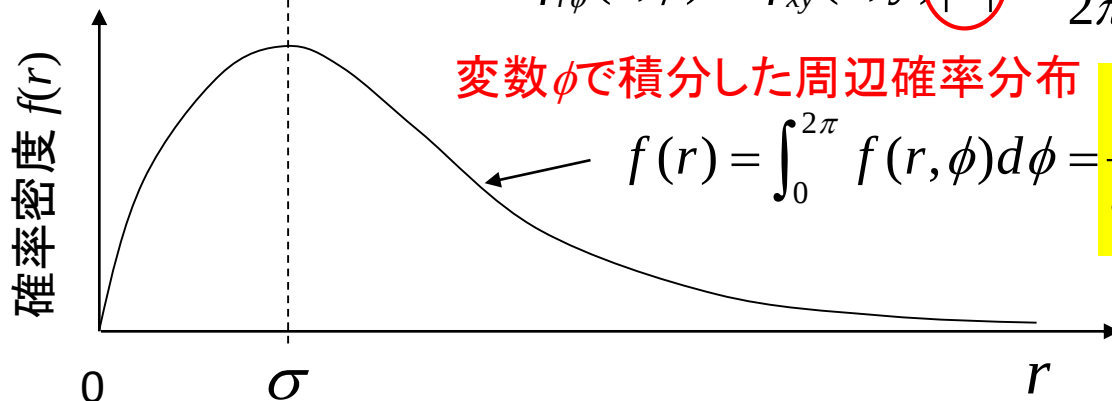
$$f_{xy}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

ヤコビアン
の
絶対値

レイリー分布

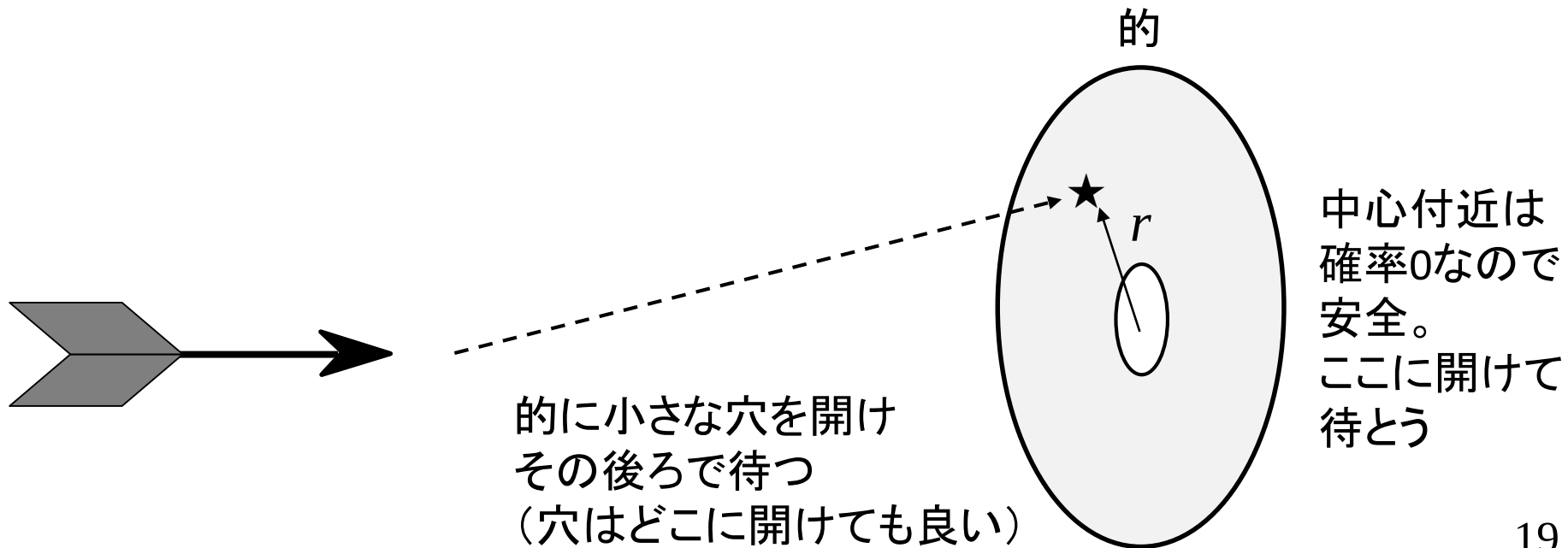
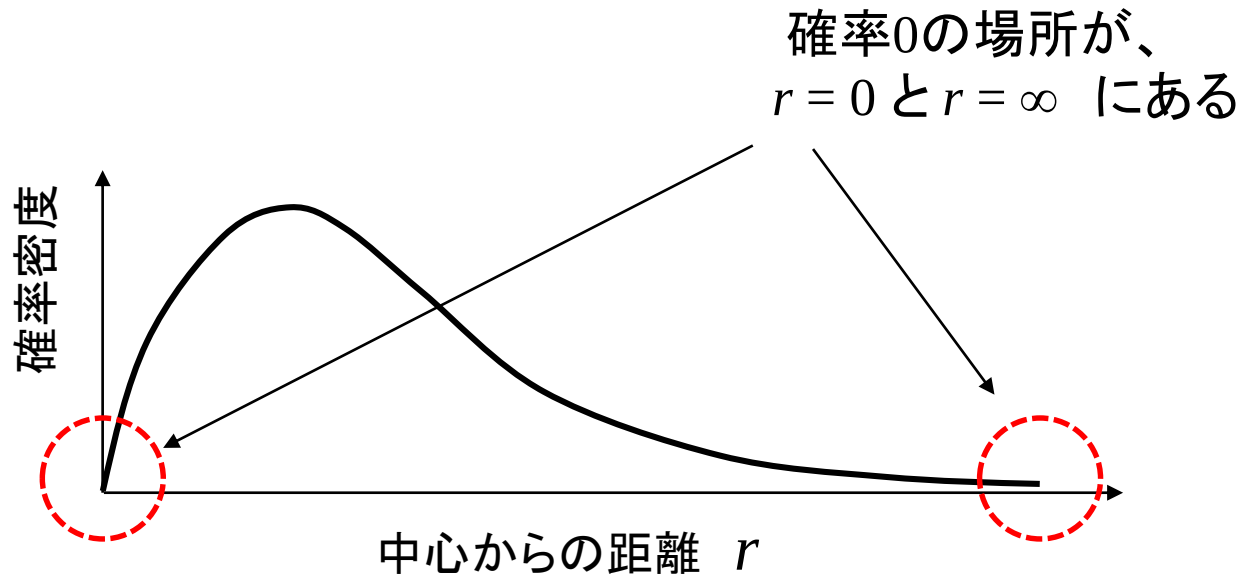
$$f_{r\phi}(r, \phi) = f_{xy}(x, y) |J| = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

変数 ϕ で積分した周辺確率分布



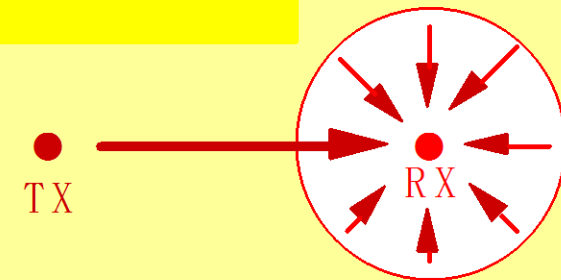
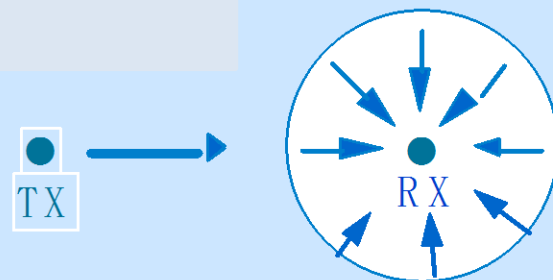
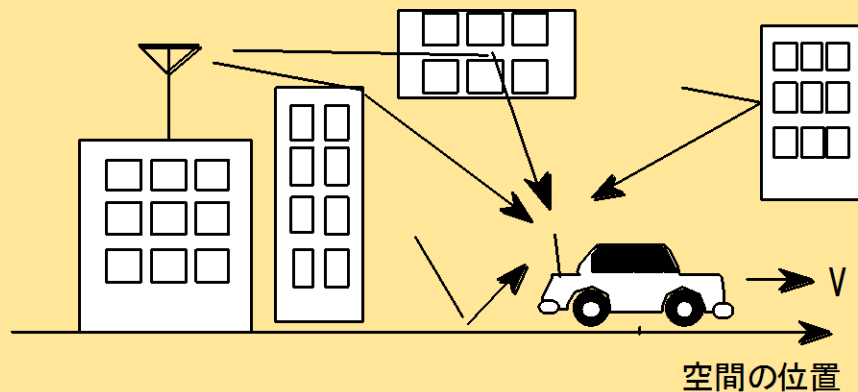
$$f(r) = \int_0^{2\pi} f(r, \phi) d\phi = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

レイリー分布についての勘違い



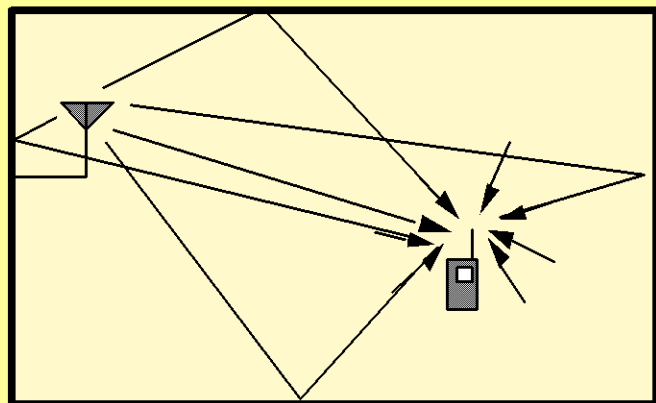
レイリーフェージング環境と仲上・ライスフェージング環境

レイリーフェージング環境

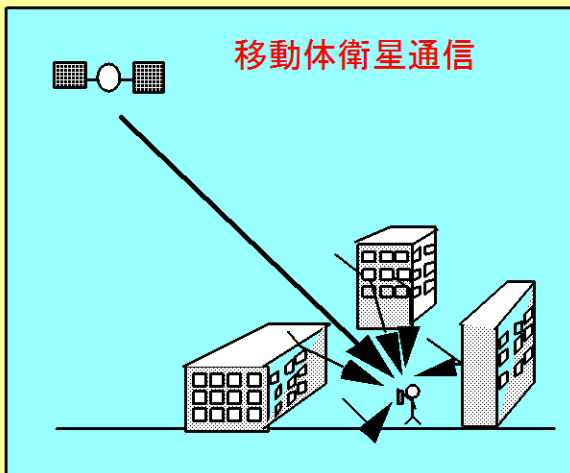


仲上・ライスフェージング環境

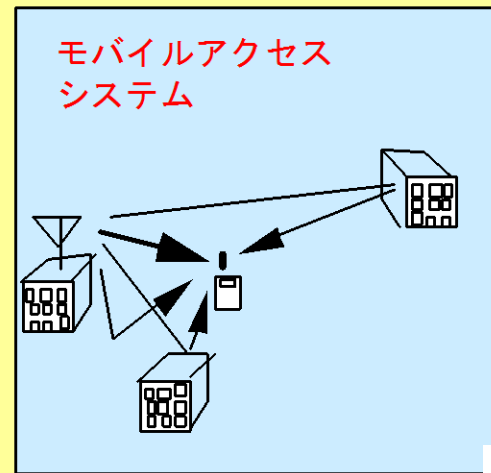
屋内通信



移動体衛星通信



モバイルアクセスシステム



② 仲上・ライス分布

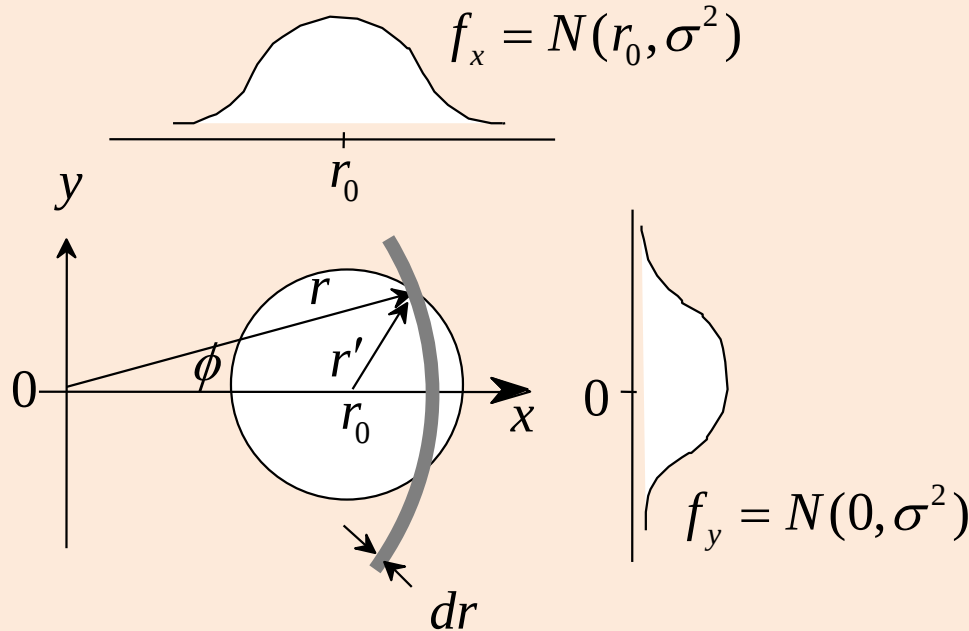
仲上稔 と S. O. Rice が独立に編み出した確率分布

見通し内フェージング環境を表すモデルに使われる

② 仲上・ライス分布

(定常成分と不規則成分の合成信号振幅の確率分布)
(仲上: 1940、Rice: 1945)

変数 x が $N(r_0, \sigma^2)$, y が $N(0, \sigma^2)$ であるときの $x+jy$ の振幅 r の分布



$$r = |x + jy| \left(= \sqrt{x^2 + y^2} \right)$$

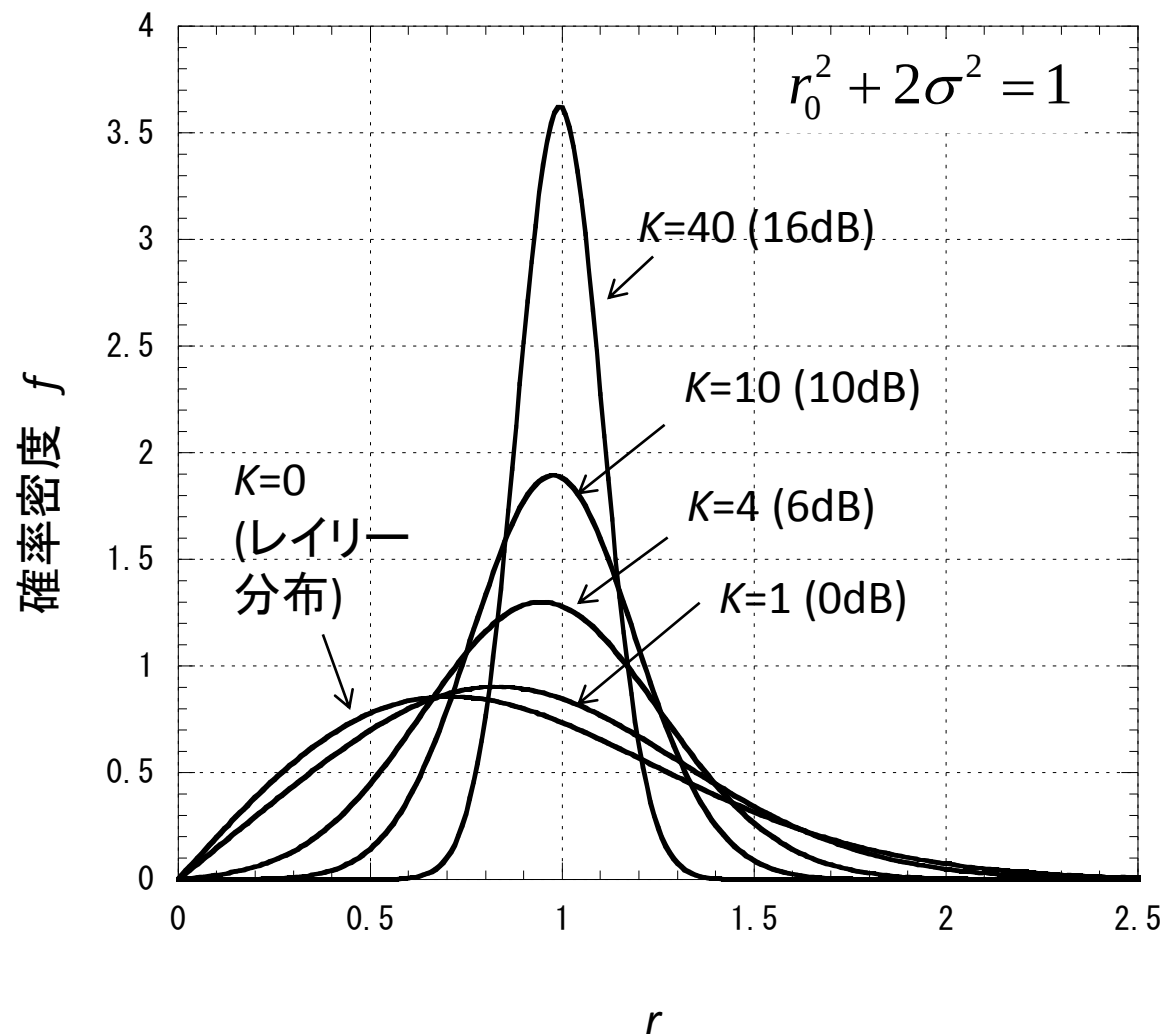
$$\begin{aligned} f_{r\phi}(r, \phi) &= f_x(x) f_y(y) |J| \\ &= \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r_0^2 - 2r_0r \cos \phi + r^2}{2\sigma^2}\right) \\ &= \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r_0^2 + r^2}{2\sigma^2}\right) \exp\left(\frac{r_0r \cos \phi}{\sigma^2}\right) \end{aligned}$$

r : 仲上・ライス分布
 r' : レイリー分布
($r_0=0$ で両者は一致)

$$\begin{aligned} f_r(r) &= \int_0^{2\pi} f_{r\phi}(r, \phi) d\phi \\ &= \frac{r}{\sigma^2} \exp\left(-\frac{r_0^2 + r^2}{2\sigma^2}\right) I_0\left(\frac{r_0r}{\sigma^2}\right) \end{aligned}$$

(変形ベッセル関数 I_0 の積分表示公式より) 22

仲上・ライス分布の計算例



ライスファクタ: K
(定常成分と
不規則成分の電力比)

$$K \equiv \frac{r_0^2}{2\sigma^2}$$

見通し内(LOS)伝搬環境
を表す

仲上稔氏が研究開発に従事したころの時代背景

1930年代後半～1940年代

（昭和の激動期：太平洋戦争を挟んだその前後）

無線通信（国際通信）は短波通信の時代

（HF: 3～30MHz, 電離層伝搬）

所属： [国際電気通信株式会社](#)（当時の国策会社）

勤務地： 福岡受信所（現埼玉県ふじみ野市にあった短波の受信所）

伝搬データ： 記録紙やブラウン管の時代、データ処理も手作業
（現代は、デジタル機器による自動収録・自動処理の時代）

この詳細は、唐沢研究室ホームページより公開している技術レポート
TR-YK-076「[仲上三分布への歴史探訪](#)」にまとめている。

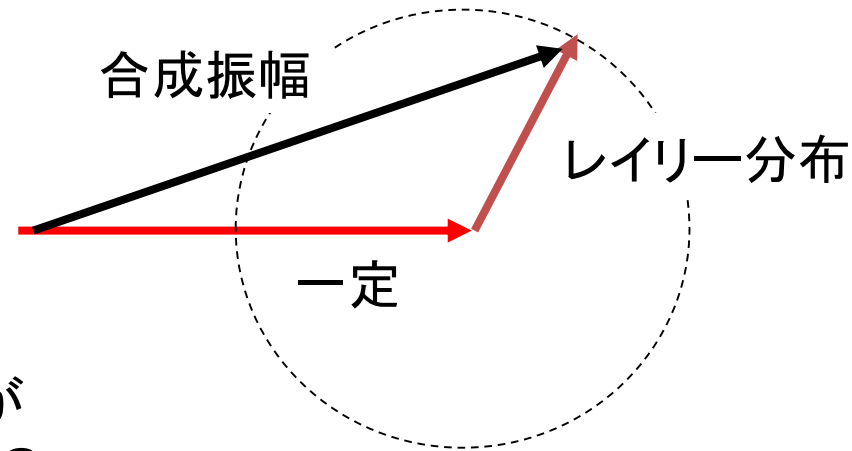
仲上・ライス分布の名前の由来

仲上稔(国際電気通信株式会社)
短波のフェージングの研究

目的; 多数の弱い電波の中に強い電波が
一つ加わるときの振幅の確率分布は?
解析手段: ハンケル変換形特性関数

S. O. Rice(米国 ベル研究所)
雑音理論の研究

目的: 一定振幅の信号に雑音加わった
ときの信号強度の確率分布は?
解析手段: 非中心2次元正規分布の振幅
確率(スライドに示した方法)



⇒ 電気通信学会誌(1940)



$$f(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r_0^2 + r^2}{2\sigma^2}\right) I_0\left(\frac{r_0 r}{\sigma^2}\right)$$

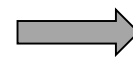


⇒ ベル研の技術論文誌
BSTJ(1944/1945)

この分布は、洋の東西の同時代の研究者が、違う目的・違う方法で編み出された。

Rice distribution / Rician distribution

ライス分布



Nakagami-Rice distribution

(我が国先人の努力) 仲上・ライス分布

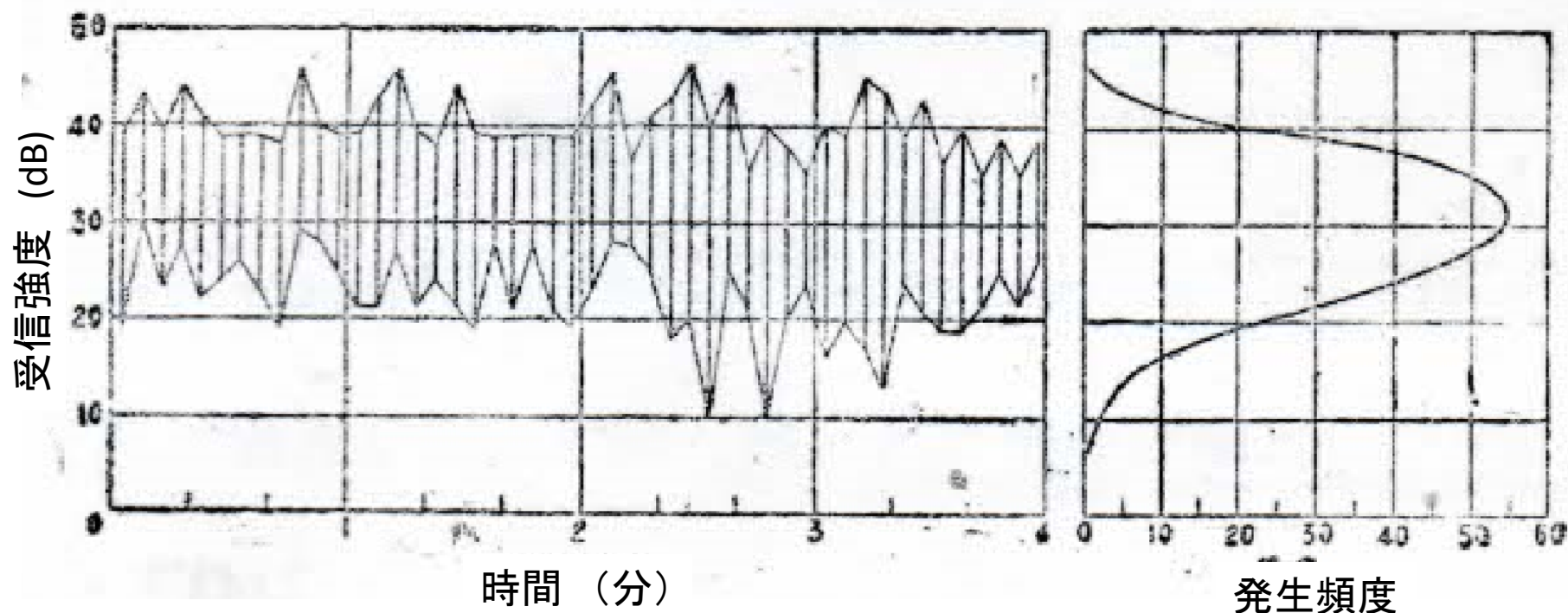
③ 仲上m分布(または、m分布、仲上分布)

移動通信のフェージング環境を表すのによく用いられる分布
世界の中で Nakagami distribution と言えば、この分布

謎に包まれた分布

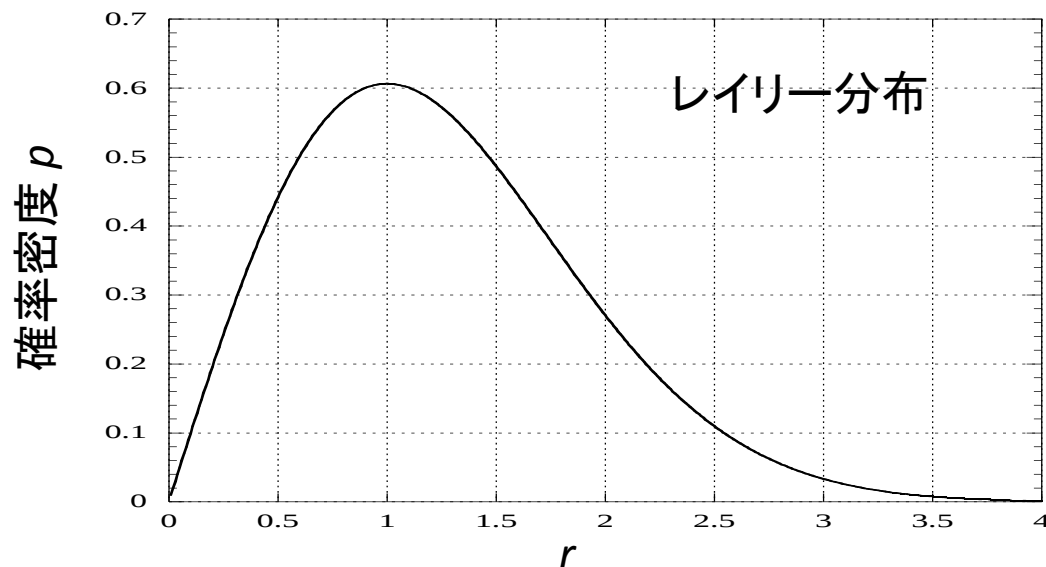
観測から生まれた経験式 → 基本的な理論分布の一つ

短波フェージングデータと頻度分布特性(下記文献の図2. 3より)



今のように自動記録したデータをパソコンが解析してくれる時代と違って、仲上が働いていた時代（1930～1940年代）には、測定データもブラウン管を見て写真を撮ったり自記記録計の紙データであったりで、解析もすべて手作業である。短波のフェージングの周期は数秒程度以下であるので、観測単位時間を3～5秒とし、その間における電界強度の変動の範囲（最大値と最小値）をもってその間の電界強度の変動を代表させ、これの数分間の連続記録（左）から頻度分布（右）を得ていたようである。

相対頻度グラフ(レイリー分布の確率密度関数を直線で表す特殊方眼紙)

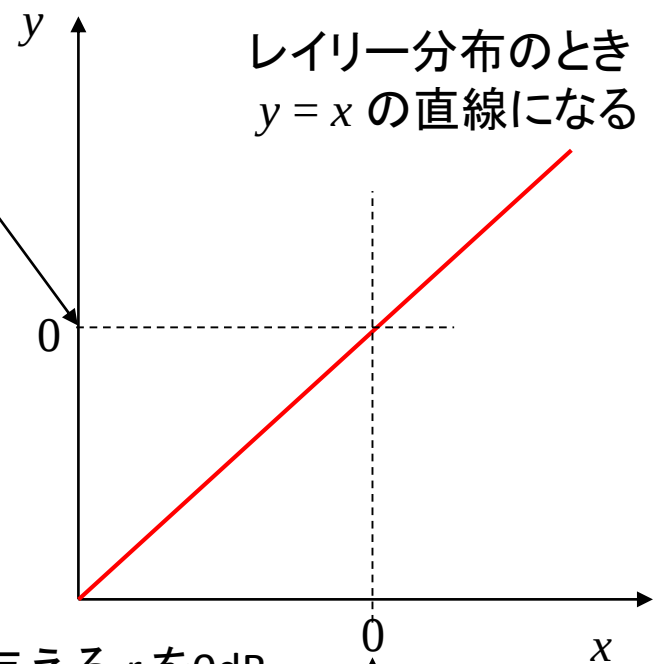
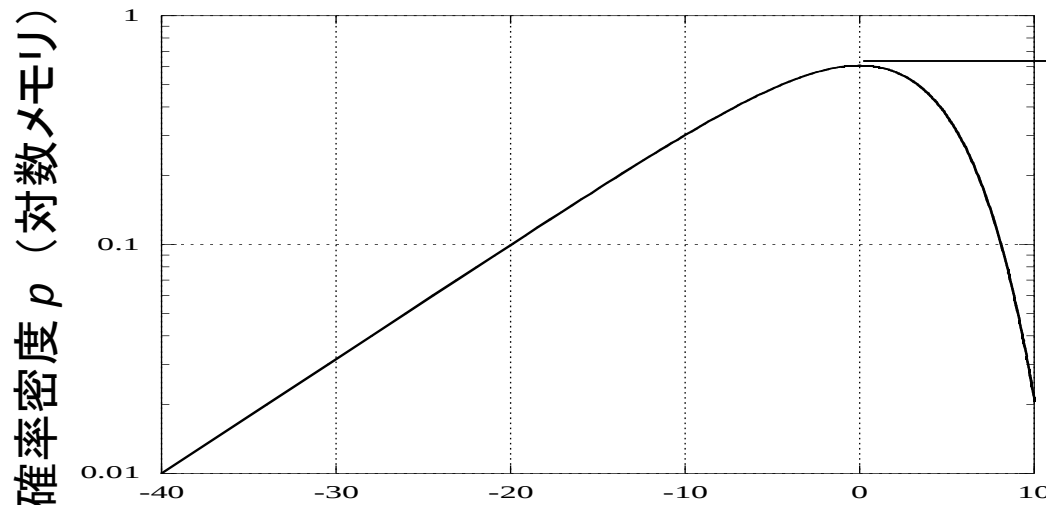


$$x = 1 + \frac{2}{M} - \exp(2\chi / M)$$

$$(M = 20 \log_{10} e)$$

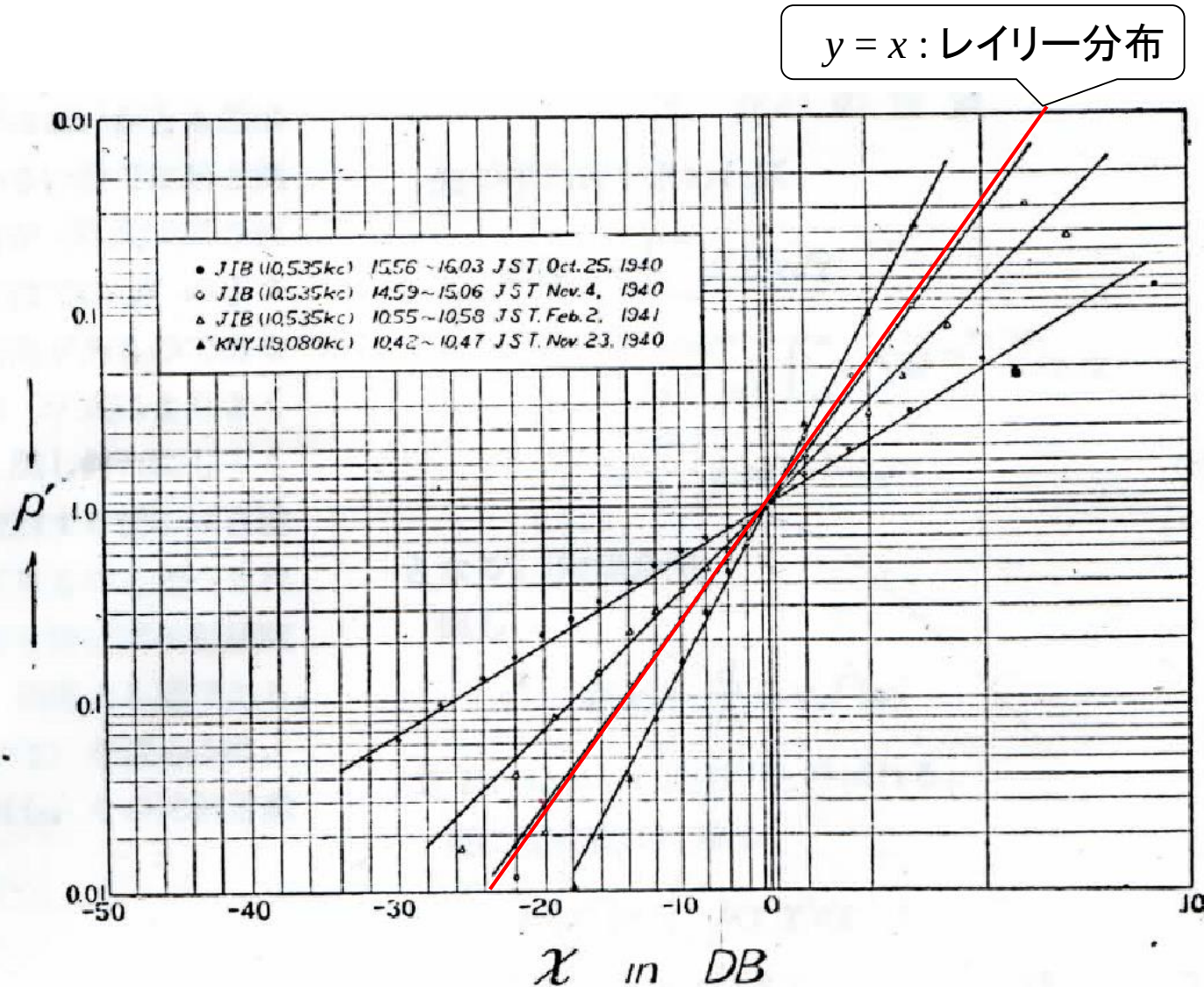
($\chi < 0$ のとき。 $\chi \geq 0$ では $-x$ に)

$$y = \begin{cases} \log_e p' & (\text{for } \chi \leq 0) \\ -\log_e p' & (\text{for } \chi > 0) \end{cases}$$



ピーク値を与える r を 0dB

短波フェージングの相対頻度特性（下記文献の図1より）



測定データから
どれも、直線に近い



$$y = mx$$

の性質を発見
(パラメータ: m)

仲上分布、m分布と呼ばれる場合もある。短波のフェージングデータから見出された分布であるが、移動通信のモデルにも役立つ非常に汎用的な分布

相対頻度グラフから $y = mx$ を発見

↓ 分布の式の導出は最後のスライドに示す資料中に

$$f(r) = \frac{2m^m}{\Omega^m \Gamma(m)} r^{2m-1} \exp\left(-\frac{m}{\Omega} r^2\right) \quad m \geq 0.5, \quad \Omega = \langle r^2 \rangle$$

↓ 式の形(見かけの複雑さに騙されないように)

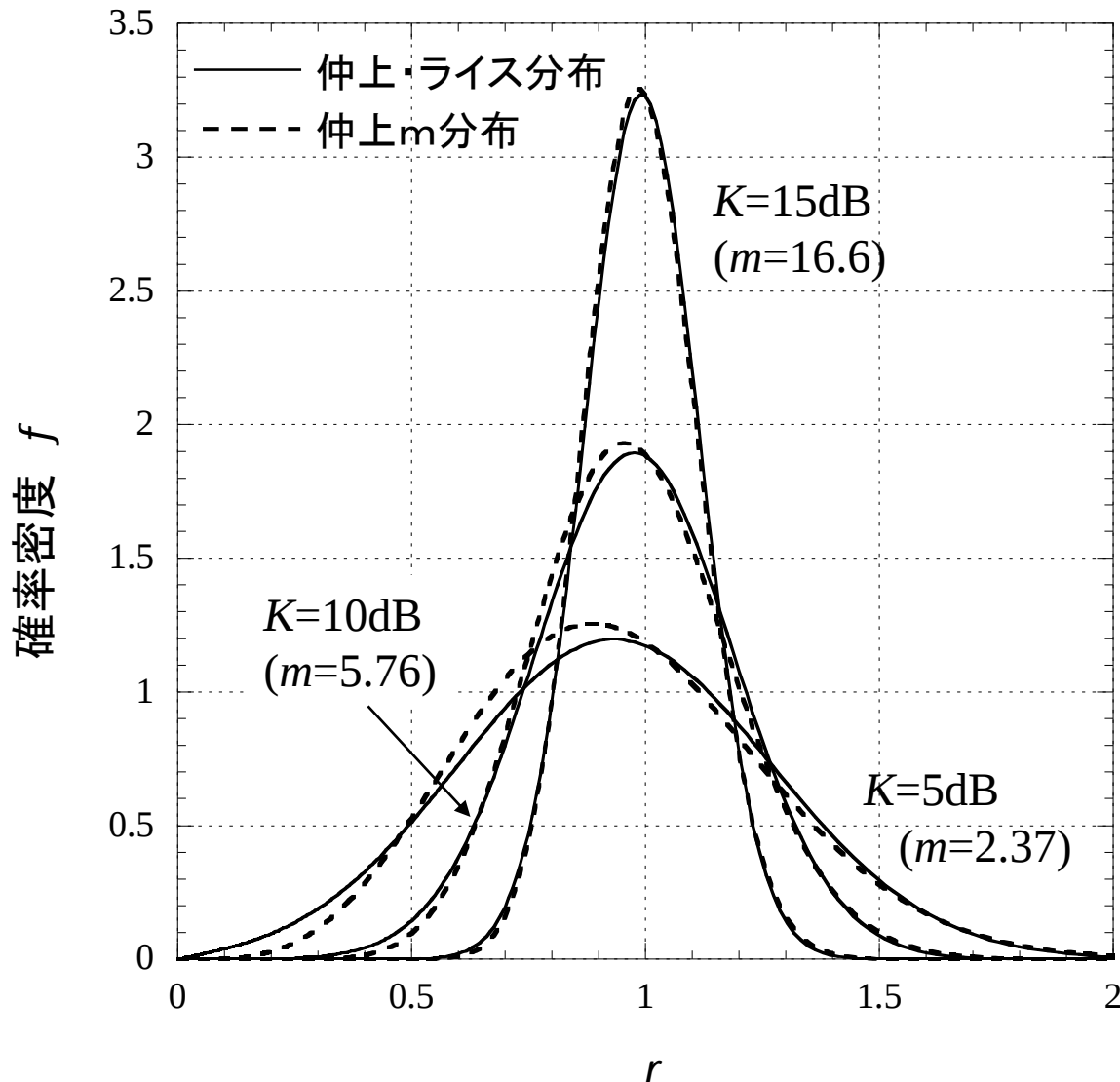
$$f(r) \propto r^{2m-1} \exp\left(-\frac{m}{\Omega} r^2\right) \quad \begin{array}{l} (m=1 \text{ でレイリー分布} \\ m=0.5 \text{ で半ガウス分布}) \end{array}$$

仲上・ライス分布と分布の形が近い(近似関係にある)。

他と組み合わせての利用する場合に、仲上・ライス分布よりは、解析性に優れている。

仲上・ライス分布と仲上m分布の近似関係:PDFで見る

$$\langle r^2 \rangle = a^2 + 2\sigma^2 = \Omega = 1 \quad (\text{電力正規化})$$



パラメータ: m と K の換算

$$m(\geq 1) = \frac{(K+1)^2}{2K+1}$$

$$K = \sqrt{m^2 - m} + m - 1$$

全体的に見ると
良い近似関係

他の関数と組み合わせて
使うとき、解析性が良いの
で、仲上・ライス分布の代
わりにm分布が使われるこ
ともある

仲上m分布:その本質は？

測定データから経験的に見つけた分布なので、
物理的意味を見出すのに苦労する時期が続いた

その後に見えてきた分布の本質

正規分布(平均値0) 確率変数 X, X_i (i.i.d.) ($i=1, 2, \dots, M$)

$$R = |X|$$

$$R = \sqrt{X_1^2 + X_2^2}$$

$$R = \sqrt{X_1^2 + X_2^2 + \dots + X_M^2}$$

確率分布

半ガウス分布

レイリー分布

仲上m分布

$$m = 1/2$$

$$m = 1$$

$$m = M / 2$$

(多次元正規分布の振幅、ランダムウォーク、スペースダイバーシチに現れる)

この理論的基本分布が、数学者から生まれず、電波技術者から生まれたところに
歴史の面白さを感じる

第2部 統計的推定

「近年、雨の降り方は変わってきたか」について、調べてみよう

強い雨は、ミリ波など10GHz以上の高い周波数の電波を弱くしてしまうので、無線通信においても無関心ではいられない問題

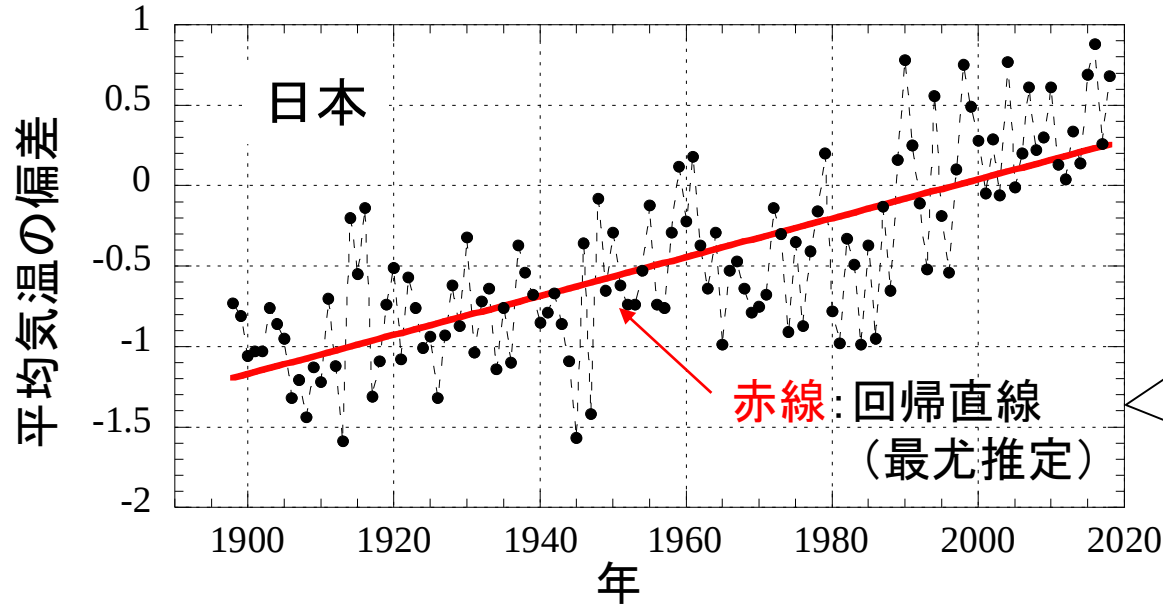
地球温暖化により、地球規模でも日本においても
この100年、徐々に気温が上昇している

では、雨の降り方にも地球温暖化の影響が現れているのだろうか？

近年、異常気象、未曾有の豪雨という言葉をよく聞くようになっている。
本当に雨の降り方は変わってきているのだろうか

気象庁が公開している日本各地の降雨データを統計的に見てみよう

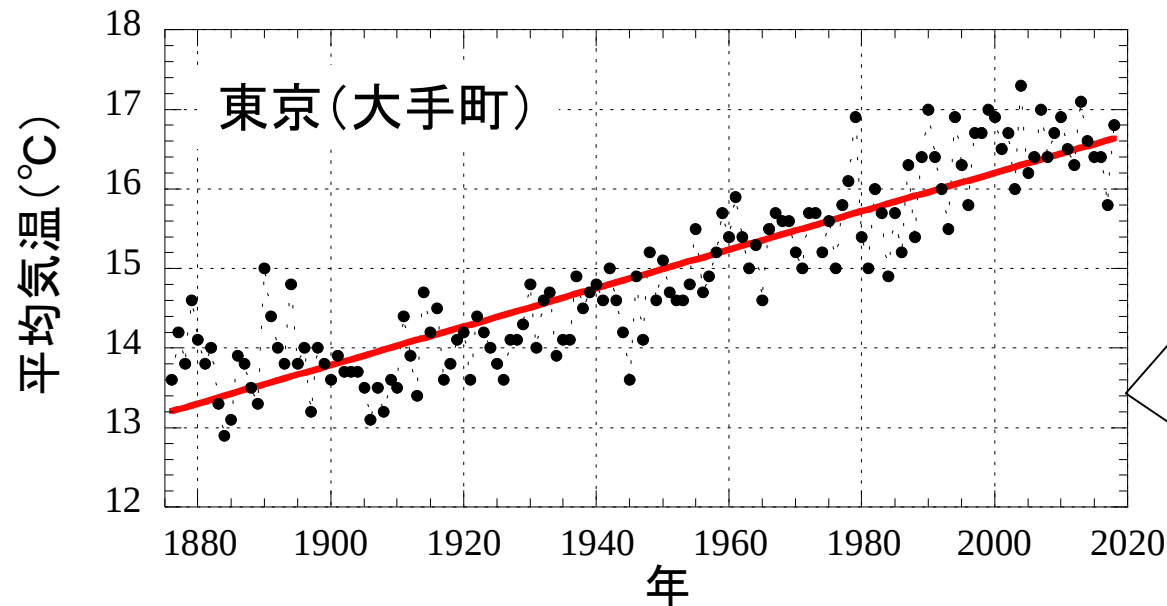
地球温暖化問題:それは確実にある



世界(地上+海上)
 $0.73^{\circ}\text{C}/100\text{年}$

日本: 都市化による影響
が比較的小さい15地点の
平均

$1.21^{\circ}\text{C}/100\text{年}$



東京(大手町にある東京管
区気象台エリア)は都市化
率(92.9%)、気温の上昇率
ともに、日本の都市の中で
最大であり、ヒートアイラン
ド現象が最も強く現れてい
るエリア

約 $3^{\circ}\text{C}/100\text{年}$

気温は、確実に年々上昇しつつある(地球温暖化)

では、雨の降り方も変わってきているのだろうか？

気温が上がると大気中に含まれる水蒸気量の最大値(飽和水蒸気量)も増えるので、潜在的な意味で、大雨になる可能性は高まる

でも、結論を出すには微妙かつ難しい問題

これを統計によって調べてみよう

日本各地での強い雨の降り方の長期的な変化を調べる

日降水量(1日の降水量)の年間最大値

→ 大きくなると、豪雨災害をもたらす

時間降水量(1時間の降水量)の年間最大値

→ 大きくなると、電波の減衰(強度低下)をもたらす

利用する統計的手法

- ① 回帰直線と区間推定: 間違った結論を出さないために
- ② 赤池情報量基準(AIC): より良いモデルを見出すために
- ③ 極値統計: 想定外の出来事を想定外としないために

どれも、数学的に奥の深い理論なので、ここでは、働きを知ることにとどめ、その手法が使いたいときには、最後に紹介するテキストで学んでほしい

この解析に用いた降水量データ(気象庁のHPより)

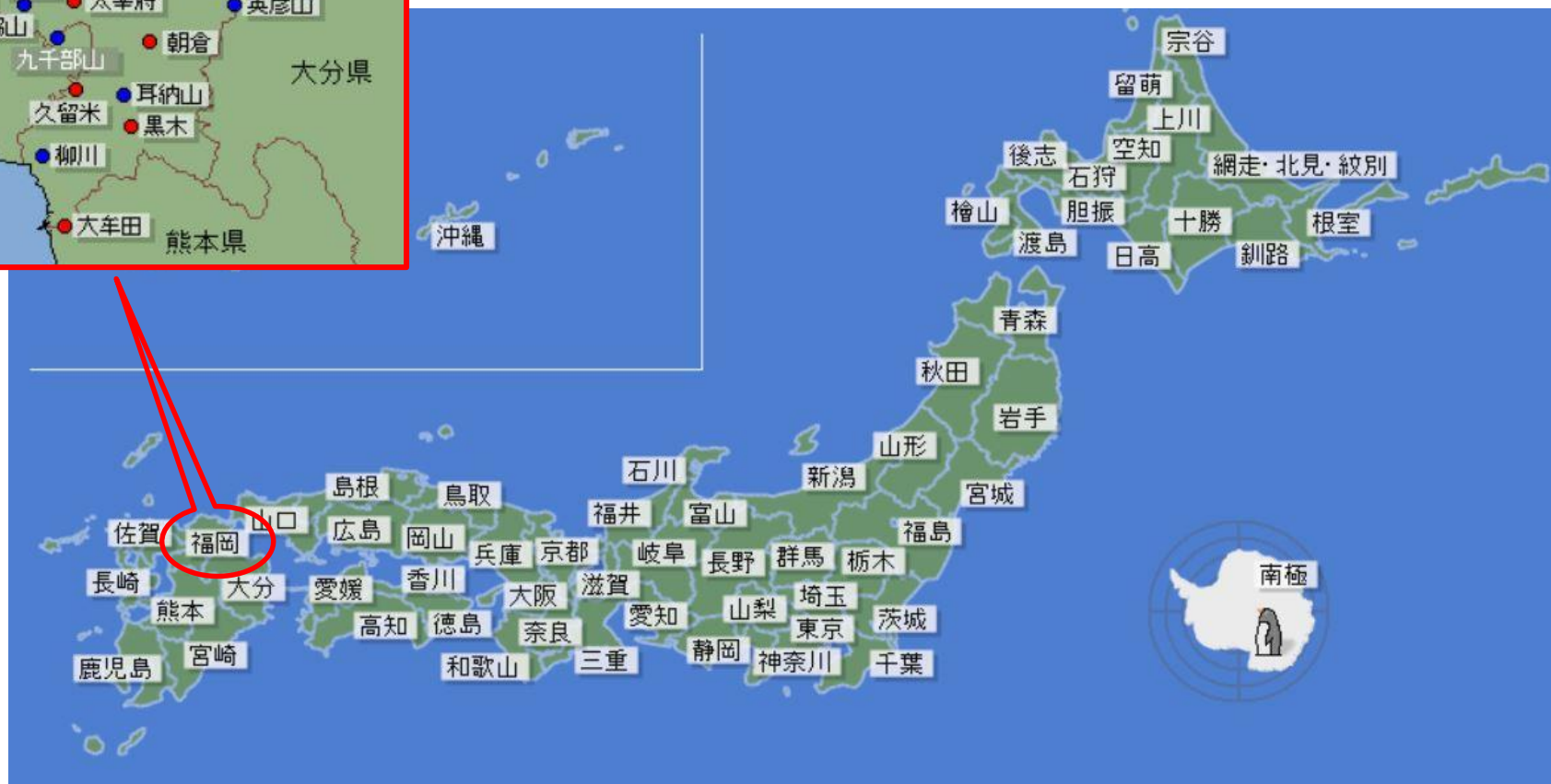
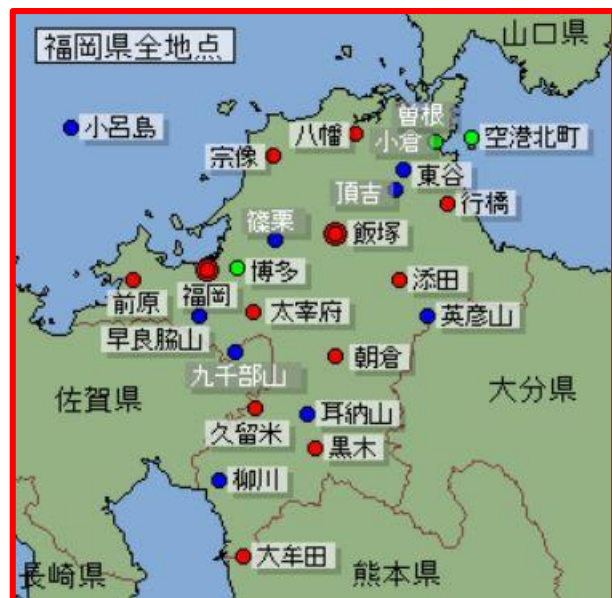
日本各地の降水量年間最大値

日降水量	45地点	100年間	(1924 - 2023)
	45地点	130年間	(1894 - 2023)
時間降水量	45地点	80年間	(1944 - 2023)

降水量の年間最大値は全て降雨によるものなので
降雨量と読み替えられる

日本の気象観測地点(約1300地点) (観測データは気象庁の ホームページから公開)

(1976年以降は主にAMeDAS)



年ごとの値

一覧表

グラフ

主要要素

詳細(気圧・降水量)

詳細(気温・蒸気圧・湿度)

詳細(風)

詳細(日照・雪・その他)

地点

降水量(最大)

東京 年ごとの値 主要要素

年	気圧(hPa)		合計	降水量(mm)			気温(℃)					湿度(%)		平均 風速	最 風速
	現地 平均	海面 平均		最大			平均			最高	最低	平均	最小		
				日	1時間	10分間	日平均	日最高	日最低						
1875			1219.2]	94.0]			17.0]	21.9]	12.2]	35.1]	-3.6]				
1876		1014.7	1755.5	76.2			13.6	18.7	8.3	35.6	-9.2	78			
1877		1015.2	1317.3	79.9			14.2	19.1	8.9	34.9	-4.8	77			
1878		1015.2	1764.2	150.9			13.8	18.3	9.6	35.1	-7.6	79			
1879		1014.5	1492.7	99.8			14.6	19.4	10.0	33.9	-5.5	77			
1880		1015.0	1685.7	116.7			14.1	19.2	9.1	33.2	-6.8	76			
1881		1015.1	1444.4	86.4			13.8	18.8	8.5	34.2	-8.4	78			
1882		1015.2	1478.3	109.7			14.0	18.6	9.0	34.2	-6.3	77			
1883		1015.0	1552.6	125.6			13.3	17.8	9.1	32.8	-7.8	76			
1884		1014.9	1314.8	72.2			12.9	17.7	8.6	33.3	-7.7	76			
1885			1531.7	149.5			13.1	17.5	9.1	31.8	-9.1	77			
1886							13.9	18.7	9.9	36.6	-7.7	75			
1887							13.8	18.6	9.7	32.6	-7.9	76			
1888							13.5	18.5	9.3	32.8	-7.5	75			
1889							13.3	18.0	9.3	33.3	-8.1	74			
1890							15.0	19.4	11.0	33.9	-5.8	77			
1891							14.4	19.3	10.0	34.4	-6.7	72			
1892							14.0	18.6	10.0	34.4	-6.8	73			
1893		1014.0	1161.3	56.3	28.1]		13.8	18.8	9.4	34.4	-7.2	71			
1894		1014.4	1320.8	132.8	52.9]		14.8	19.6	10.6	35.1	-5.3	73			
1895		1013.8	1397.8	133.0	22.6]		13.8	18.7	9.8	33.0	-5.7	74			
1896		1014.4	1373.9	74.0	22.2]		14.0	18.7	10.1	34.3	-5.6	74			
1897		1014.6	1497.2	73.2	34.1]		13.2	17.7	9.2	34.0	-7.0	75			

「年」をクリックすると「月」が出る
「月」をクリックすると「日」が出る
「日」をクリックすると「時間」が出る
(ビッグデータ、日本の宝)

「年」をクリックすると「月」が出る
「月」をクリックすると「日」が出る
「日」をクリックすると「時間」が出る
(ビッグデータ、日本の宝)

点推定と区間推定

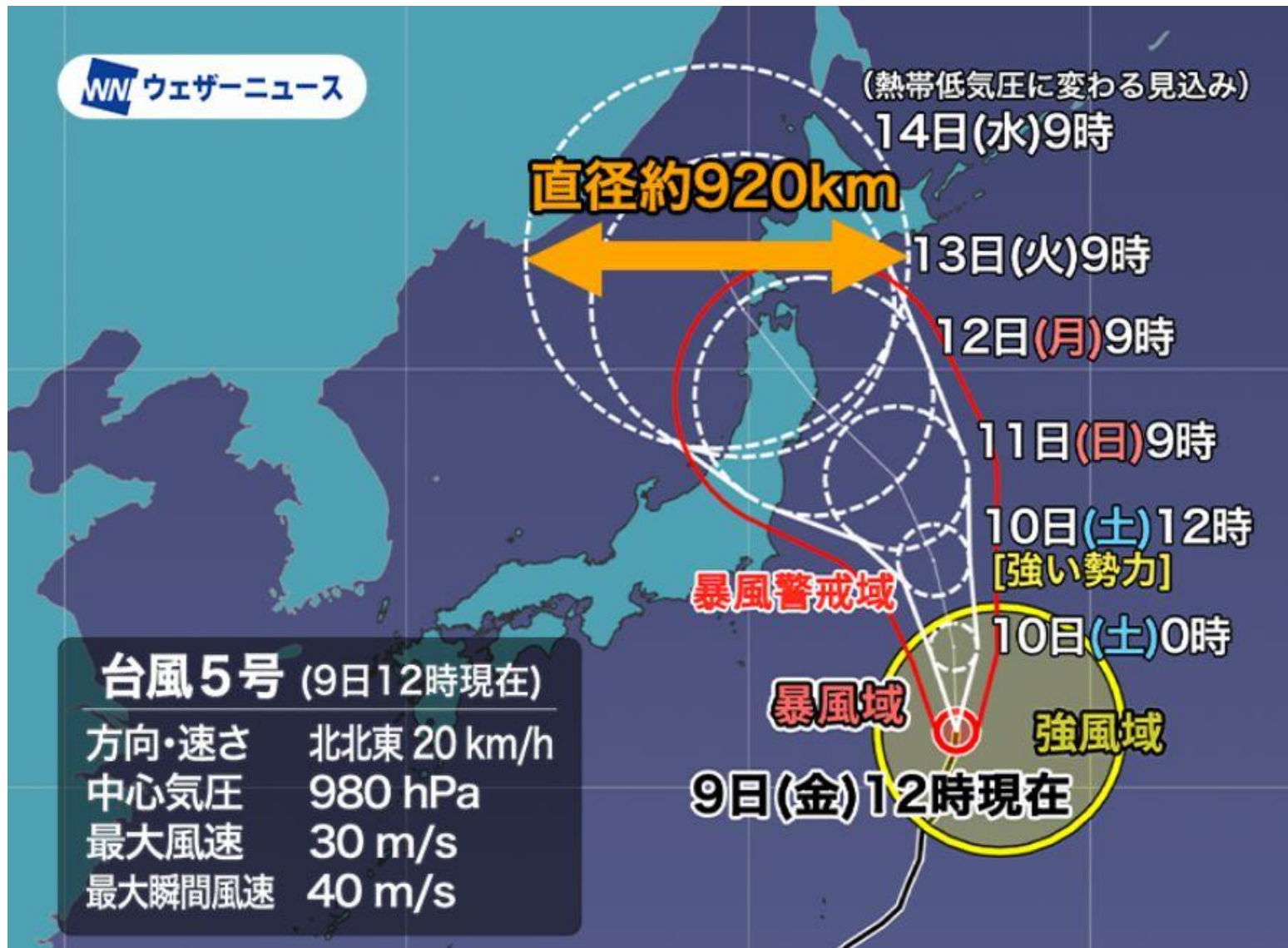
点推定: 最も確からしい値を推定する

そのようにして得られた結果(例えば、回帰直線)は、最も確からしいのであるが、それが真に正しいかどうかは分からない

区間推定: 真値が存在する可能性(確率)を幅で推定し、その幅内を**信頼区間**とする推定
(確率設定の代表値は**95%**が多い)
誤った判断をしないための用心深い考え方

点推定と区間推定の例

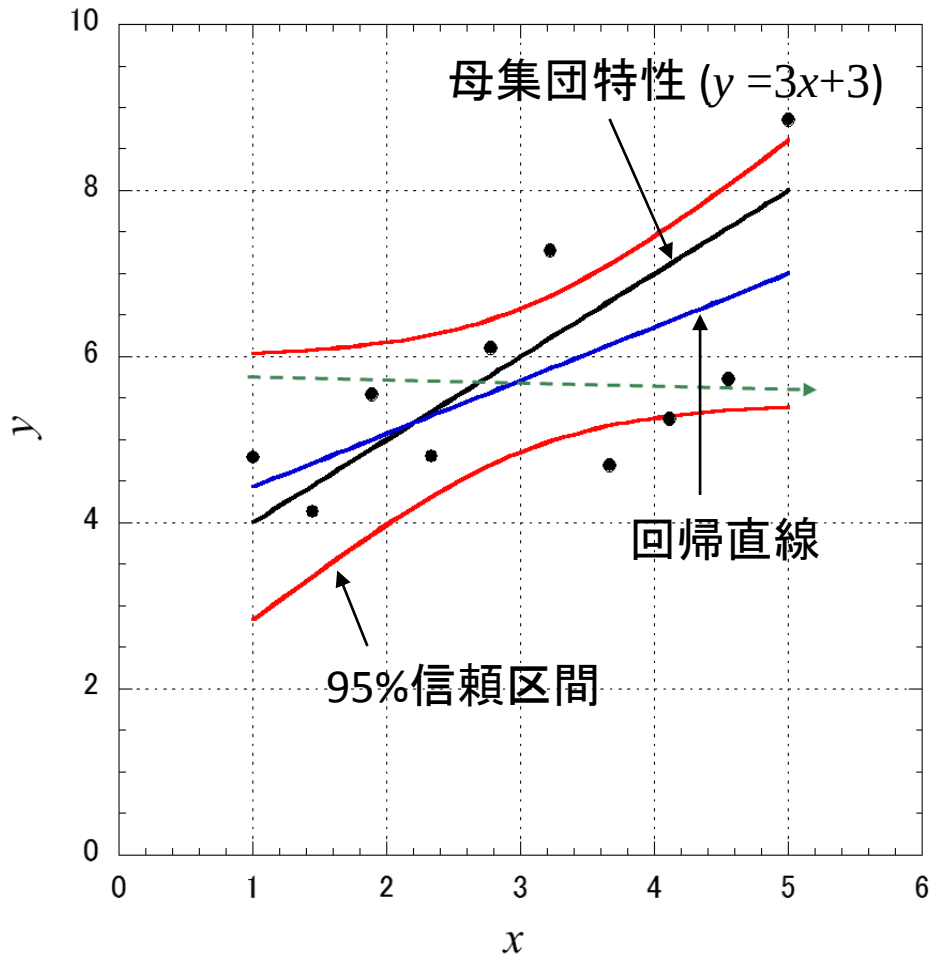
台風進路の予報円（台風の中心が70%の確率で存在する範囲）



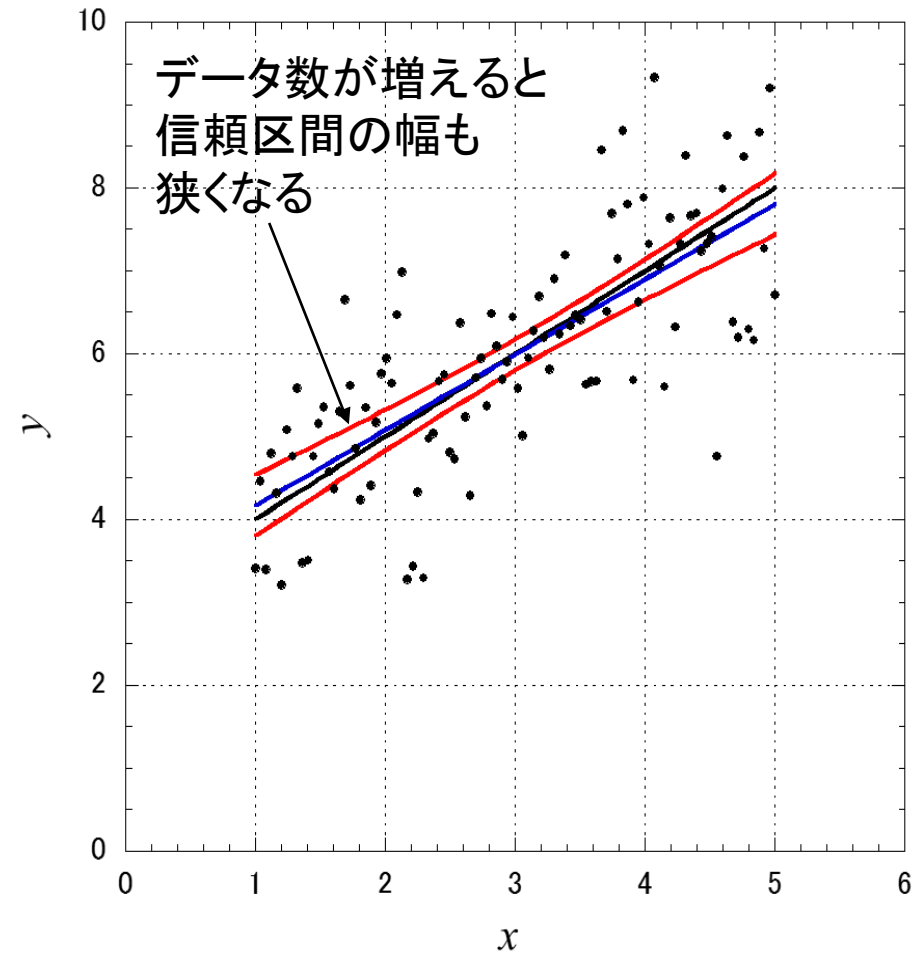
(2024年8月: ウェザーニュース画面より)

解析の道具： 直線回帰と信頼区間

データ数 10



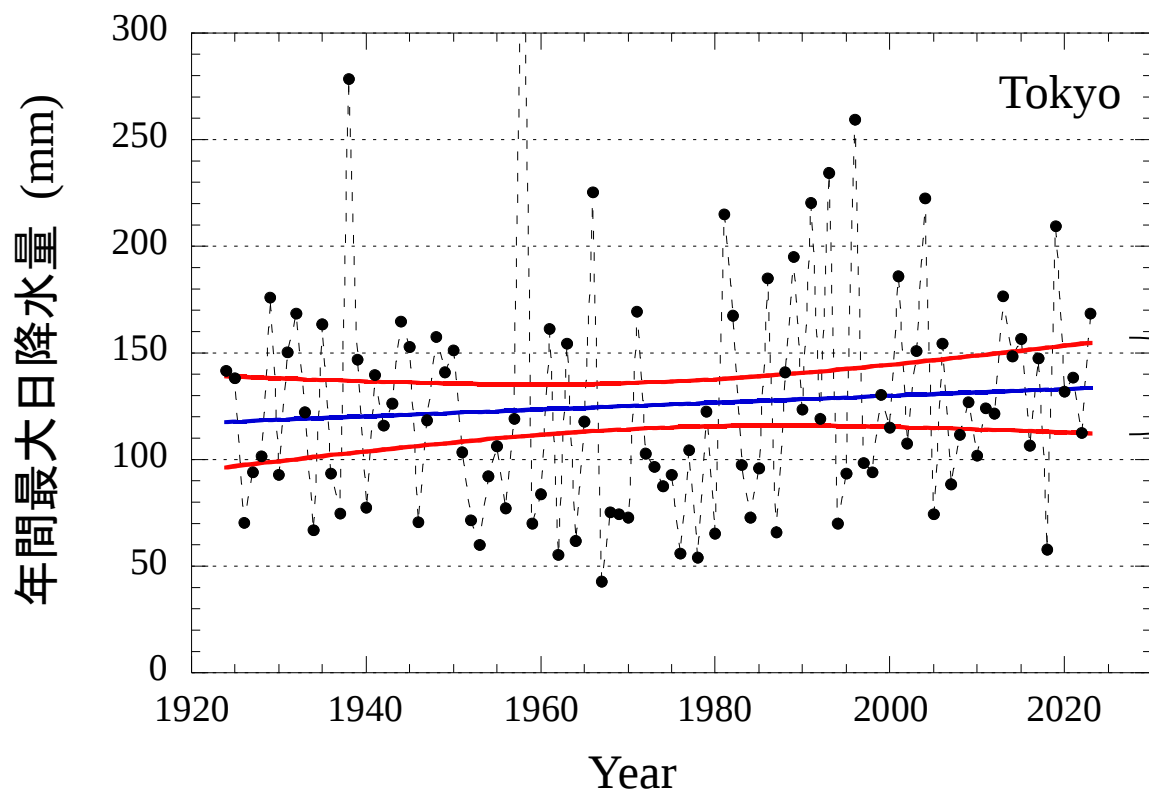
データ数 100



【信頼区間はデータのばらつき幅とはまったく違う!!】

地点別年間最大日降水量の100年間の推移

東京(千代田区)



回帰直線(青線)はやや増加の傾向を示しているが、このデータからは、「増加している」という結論は出せない

この例だけでなく、他の例(地点や降水量タイプ)を含めて、地点毎の100年間データからは、有意な長期変化傾向があるとは言えない

日本全体のデータで傾向を俯瞰的に見る

長期変化傾向を見るのに地点毎のデータではデータ数が足りない



それぞれの地点のデータをその地点の平均値で割って正規化する
(平均値を1とする年間最大正規化降水量を用いる)

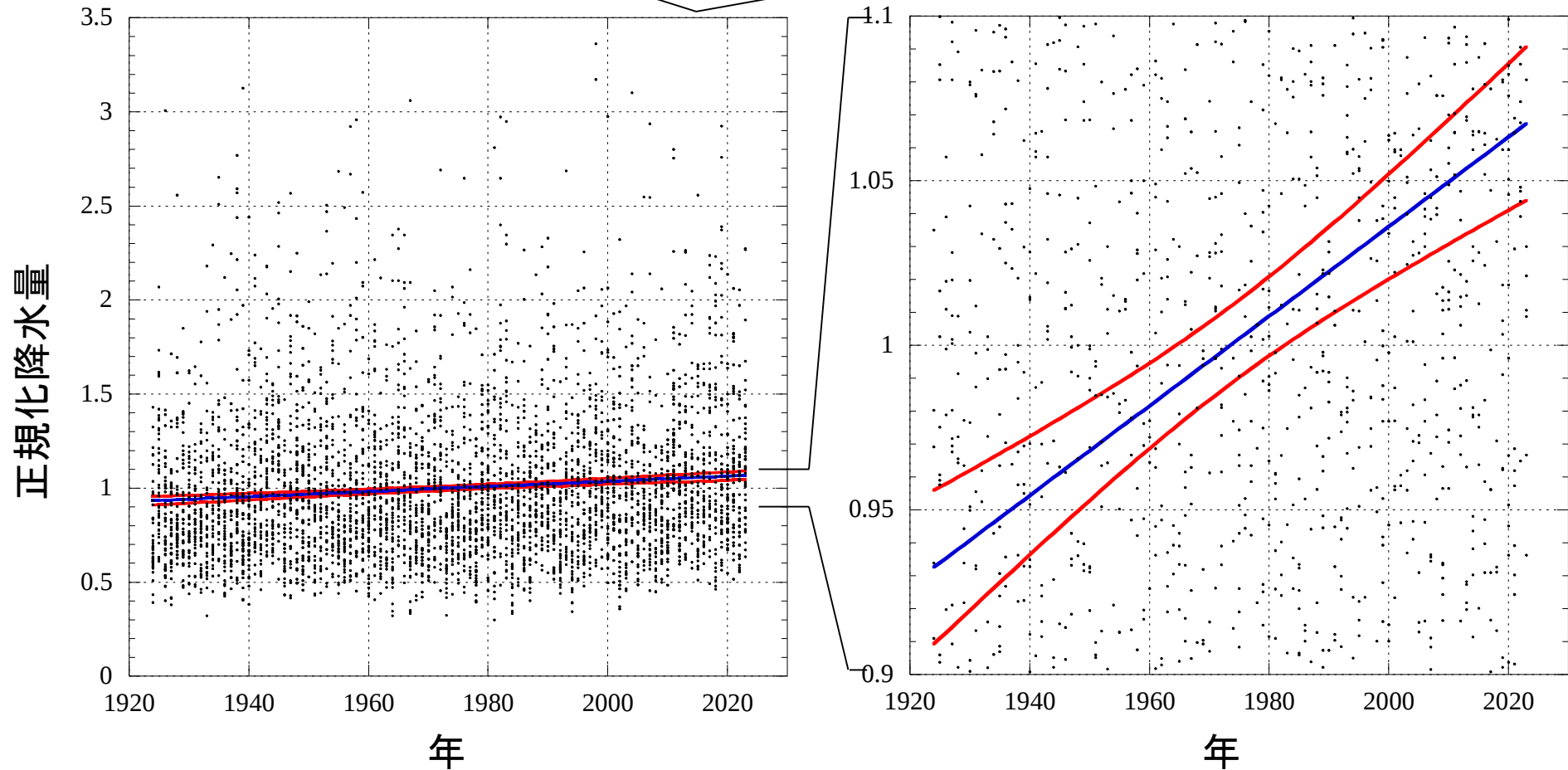
$$R_{\max, i} \rightarrow = R_{\max, i} / \langle R_{\max} \rangle$$



それらの全体のデータで、回帰直線と信頼区間を求める

日降水量

日本全国45地点の100年間分のデータ(データ点数:4500)

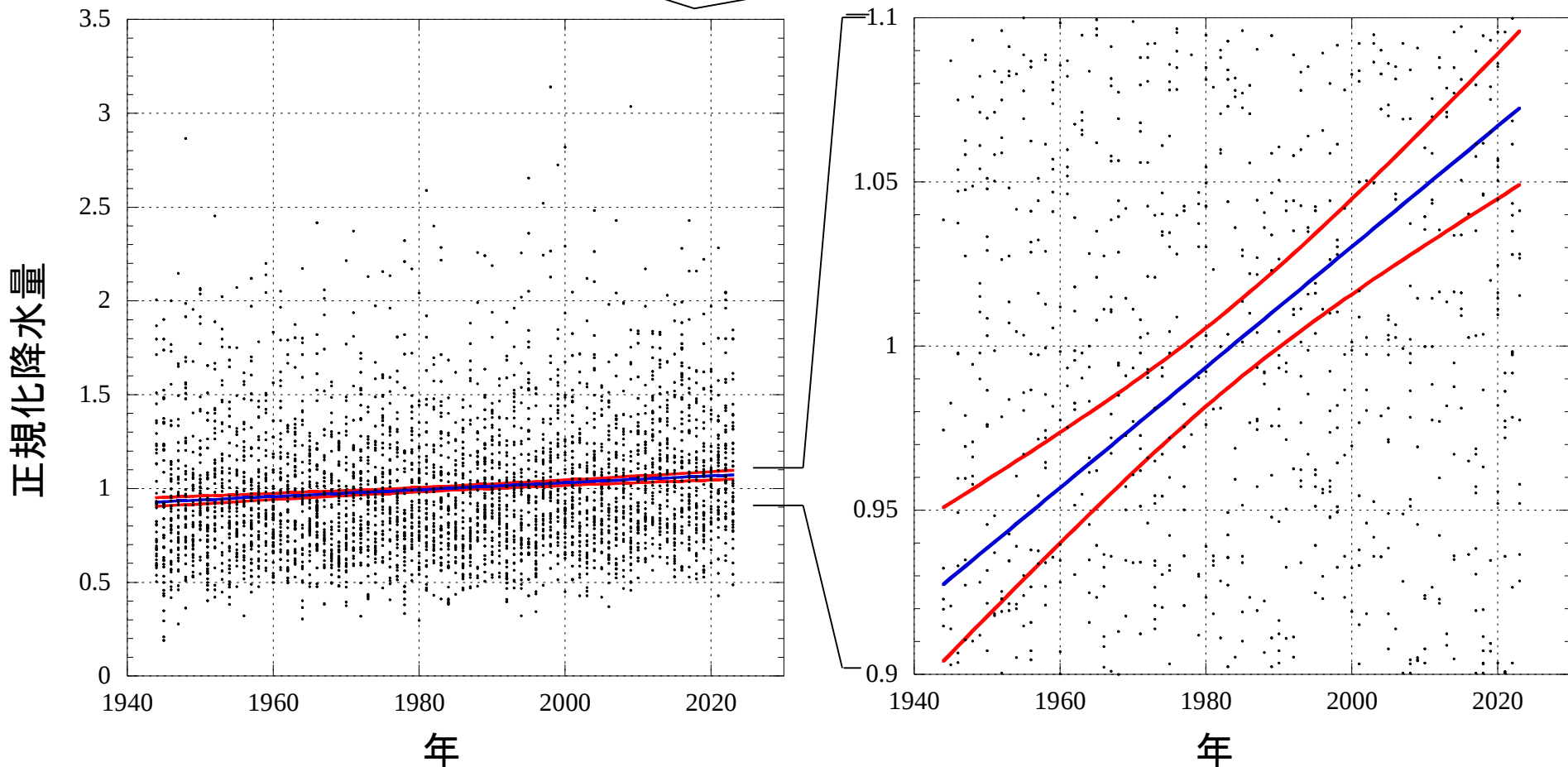


回帰直線(最尤推定): 100年間で14%の増加

区間推定(95%信頼区間): 9%~20%の増加の範囲 → 増加傾向が有意

1時間降水量

日本全国45地点の80年間分のデータ(データ点数:3600)



回帰直線(最尤推定): 80年間で16%の増加

区間推定(95%信頼区間): 10%~21%の増加の範囲 → 増加傾向が有意

長期的変化の傾向を、より詳細に分かる**回帰曲線**で見たい
どのような曲線を選ぶのが良いのだろうか？

例えば、多項式近似による最尤推定

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_Kx^K$$

多項式の次数 K を大きくすれば、推定式とデータ的一致度(＝尤度)は高くなるが、
次数の高い式を選べば良いのだろうか？

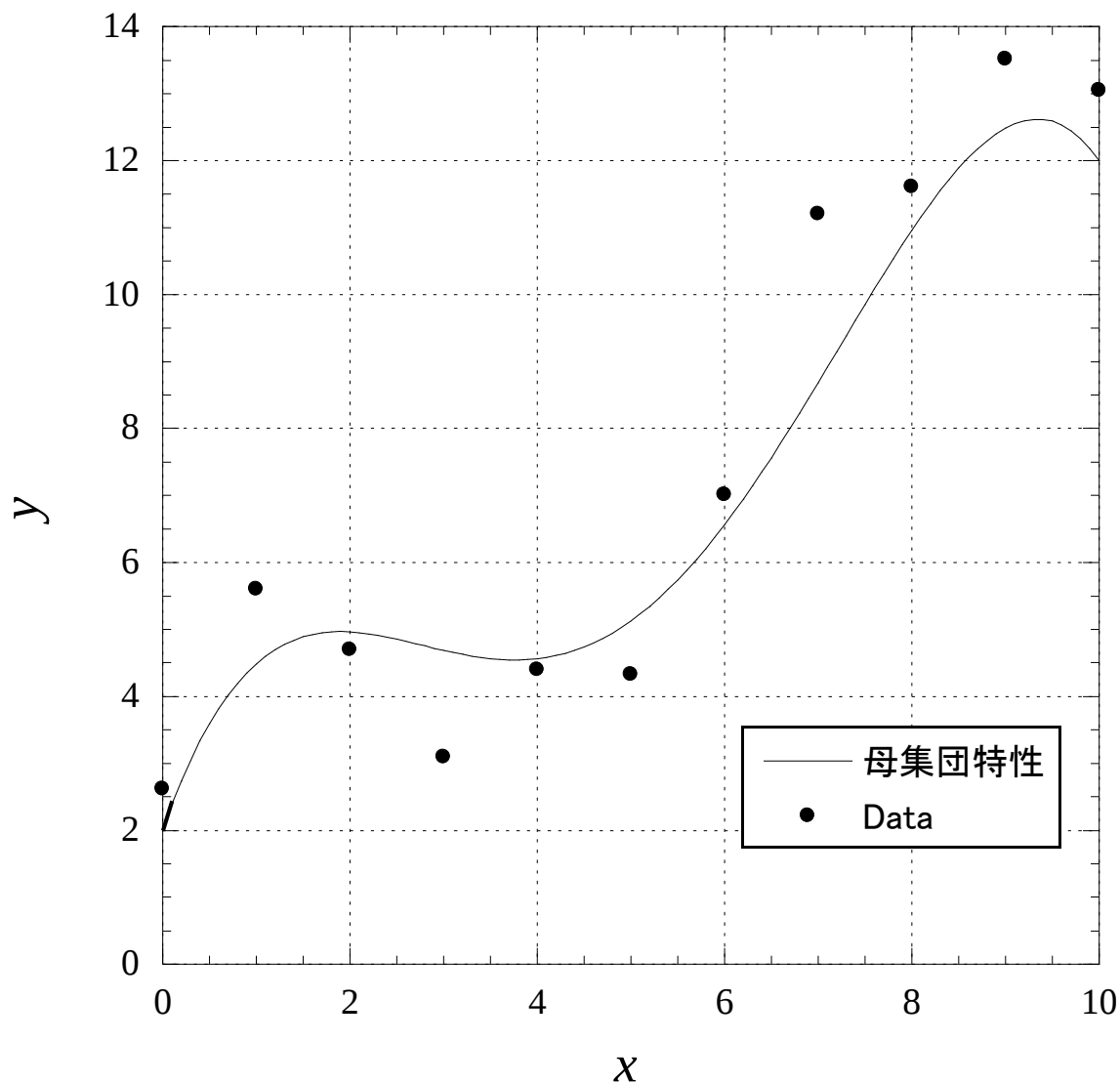
次数を上げすぎると、標本データに支配されて、他の独立なデータに合わなくなる

次数の選択には、最適値がある。

その選択基準として、**赤池情報量基準**(AIC: Akaike Information Criterion)がある

これを使ってみよう

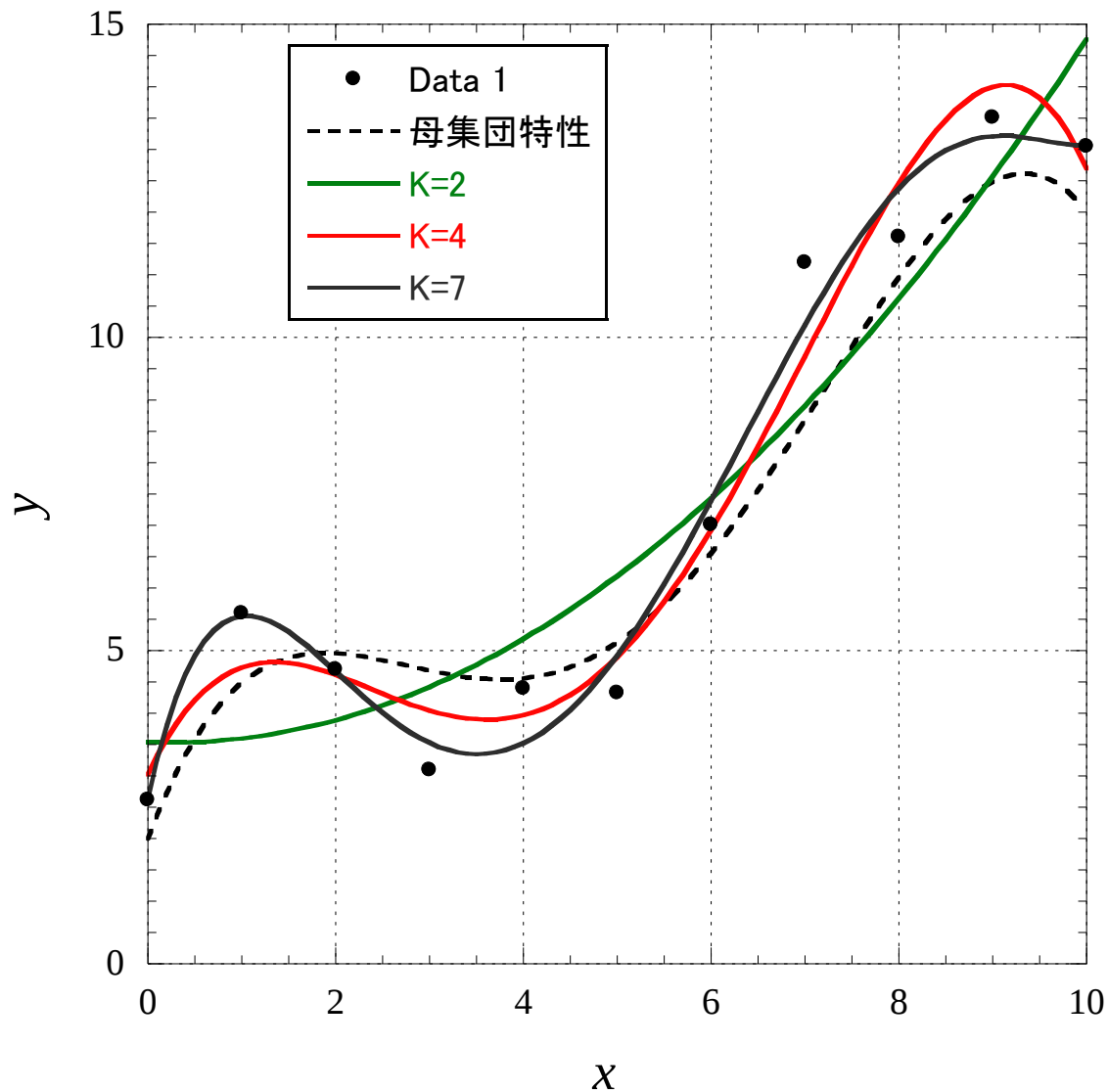
なぜ、高い次数の最尤推定では駄目なのか？



手元には、
誤差を含んだデータしかなく、
母集団の特性を知らない

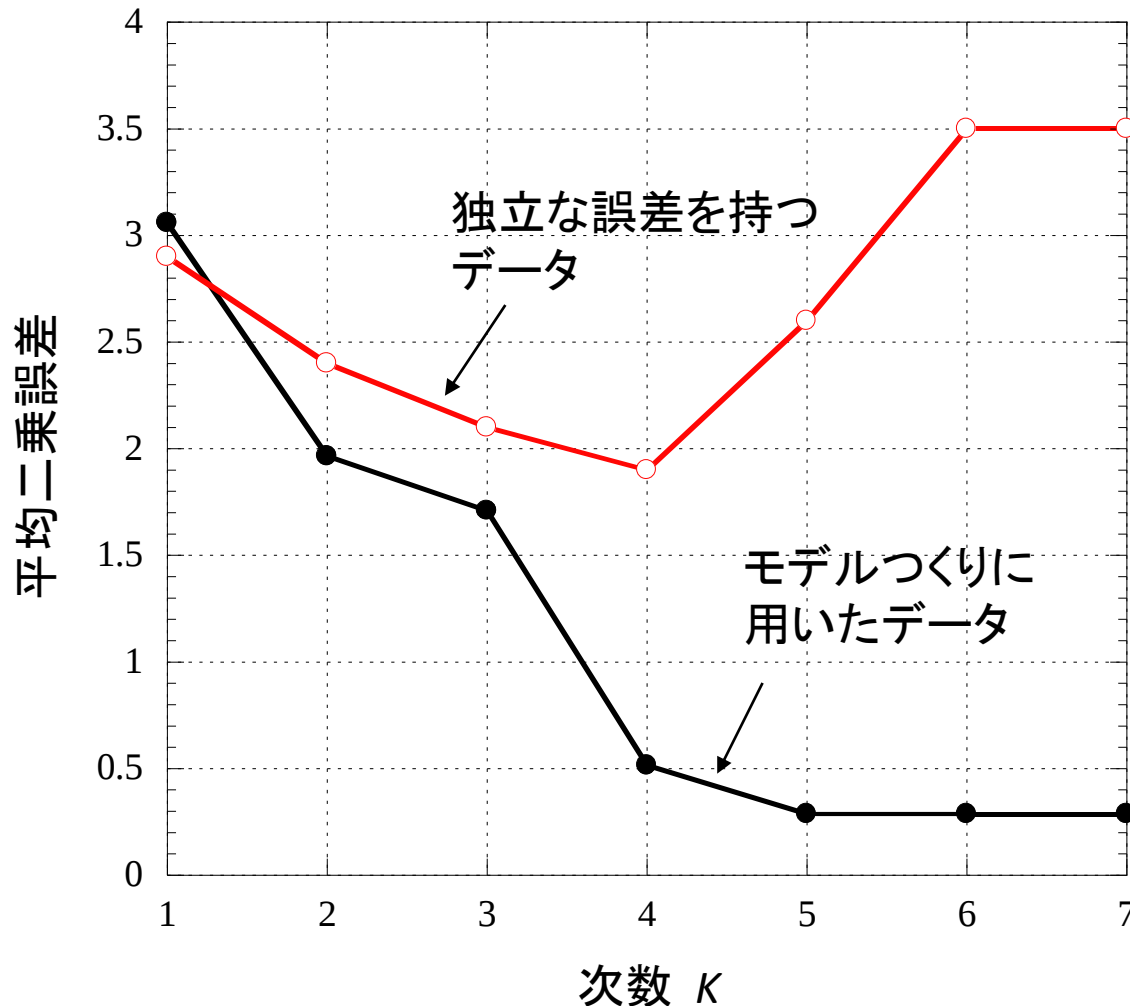
このデータから、母集団の
特性を推定してみよう

なぜ、高い次数の最尤推定では駄目なのか？

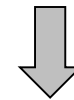


次数を上げてゆけば、
推定式との誤差は
どんどん小さくなる

なぜ、高い次数の最尤推定では駄目なのか？



次数が高くなりすぎると、独立な誤差を持つデータに対して、逆に一致度が落ちてくる



推定に際して、どこかに最適な次数がある

それを見つけるにはどうしたら良いか

最適なモデルを選ぶ指標：赤池情報量基準（AIC）

標本データから、母集団を推定する良いモデルを選ぶ判断基準

赤池弘次博士（（元）統計数理研究所：1971年考案）

良いモデルとは

- 1) モデル（推定式）と標本値の差が小さい（データによく合う）
- 2) モデルのパラメータの数が少ない

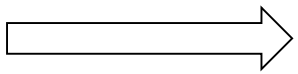
$$AIC = (-2)(\text{最大対数尤度}) + 2(\text{パラメータ数})$$

AICの値が小さいほど良いモデル

最大対数尤度とパラメータ数が同じ次元になっているのが奇跡
この基準は今でも現役で、広く使われている
理論は難しいが、確率・統計知識の宝庫

この特性は誰も知らない
(モデルの判断基準にできない)

母集団
 $g(x)$



無作為標本
 $x_1, x_2, \dots, x_n$

AICは
ここから始まる



モデルの形を決める
 $f(x; \theta)$

いろいろのモデルを
用意する



対数尤度関数を作る
$$l(\theta) = \sum_{i=1}^n \log f(x_i | \theta)$$

最尤推定値 $\max_{\theta} l(\theta) \rightarrow \hat{\theta}$

尤度が最大になるよ
うに、モデルのパラ
メータの値を決める

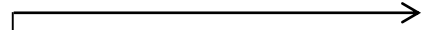


$AIC = -2 \times l(\hat{\theta}) + 2 \times K$
 $l(\hat{\theta})$: 最大対数尤度
 K : パラメータ数

AICの値を計算する

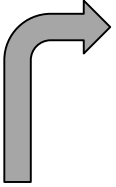
AICの値が
小さいモデ
ルを選ぶ

AIC値による
モデルの評価
選択・改良



AIC (Akaike Information Criterion: 赤池情報量規準)

AICにおけるモデルの選択基準


$$AIC = -2 \times (\text{最大対数尤度}) + 2 \times (\text{パラメータの数})$$

の式において、最も値が小さくなるモデルを選ぶ

多項式近似のモデル化(誤差が正規分布する)

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_K x_i^K + b_i \quad (f_b = N(0, \sigma^2)) \quad (\text{パラメータ数: } K+2)$$

正規分布の対数尤度関数

$$l(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \sum_{k=0}^K a_k x_i^k \right)^2 \quad \theta = (a_0, a_1, a_2, \dots, a_K, \sigma^2)$$

係数と誤差分散の最尤推定

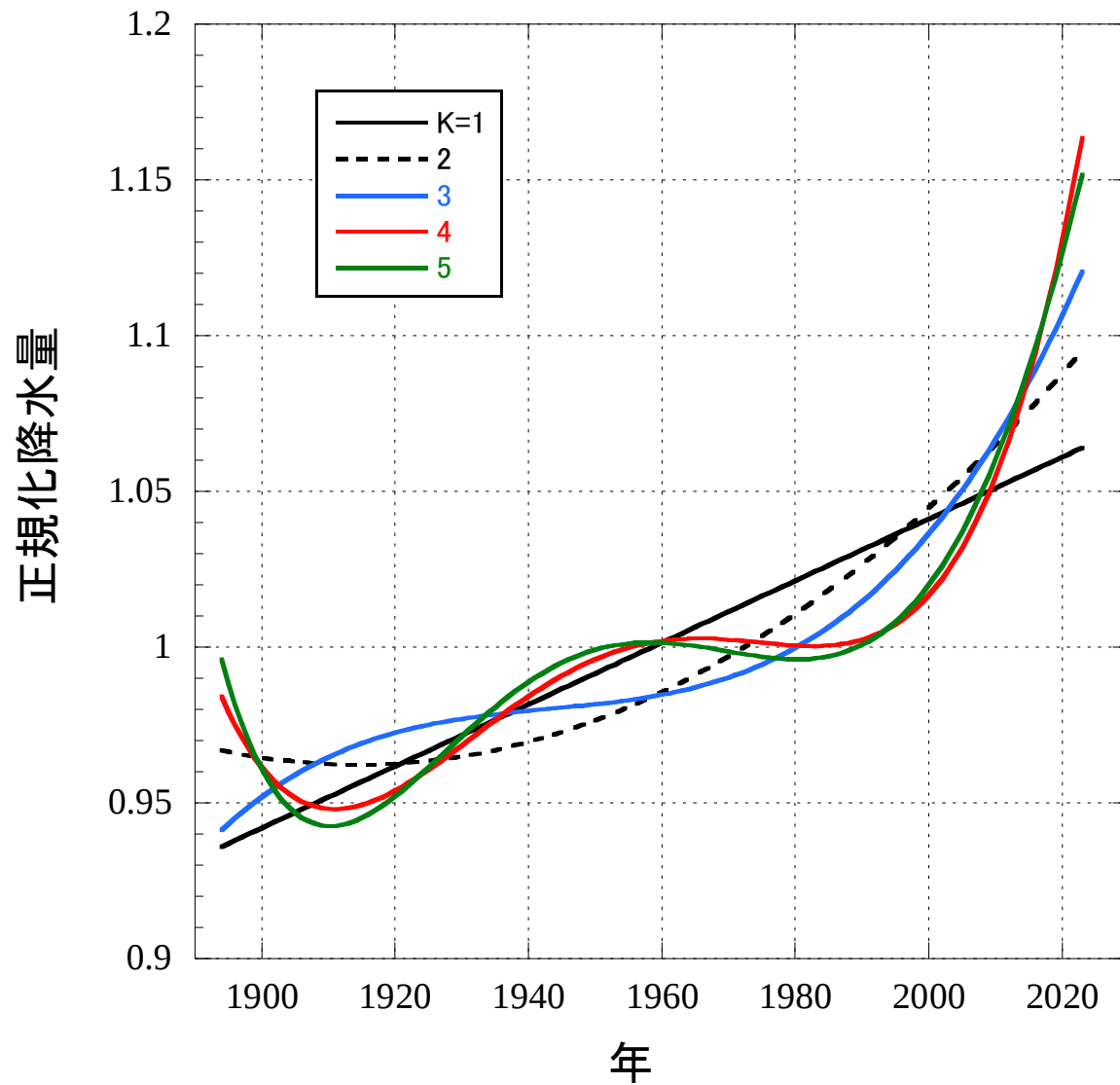
$$\frac{\partial}{\partial a_i} \sum_{i=1}^n \left(y_i - \sum_{k=0}^K a_k x_i^k \right)^2 = 0 \quad \Rightarrow \quad \hat{a}_0 \sim \hat{a}_K$$

$$\frac{\partial l(\theta(\hat{a}_0, \dots, \hat{a}_K, \sigma^2))}{\partial \sigma^2} = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{k=0}^K \hat{a}_k x_i^k \right)^2$$

最大対数尤度

$$l(\hat{\theta}) = -\frac{n}{2} \log(2\pi\hat{\sigma}^2) - \frac{n}{2}$$

日最大降水量130年間データ(多項式最尤推定)



どの次数のカーブを選んだら良いか?

多項式近似における最大対数尤度とAIC評価値

回帰曲線式 $y_i = a_0 + a_1x_i + a_2x_i^2 + \cdots + a_Kx_i^K + b_i \quad (f_b = N(0, \sigma^2))$
 (パラメータ数: $K+2$)

Rainfall	K	$\hat{\sigma}^2$	$l(\hat{\theta}^2)$	AIC	ΔAIC
1-day	1	0.16165	-2284.99	4575.99	25.03
	2	0.16160	-2284.29	4576.58	25.63
	3	0.16066	-2271.25	4552.49	1.543
	4	0.16054	-2269.48	4550.95	0
	5	0.16054	-2269.48	4552.95	1.998
1-hour	1	0.13068	-1445.22	2896.44	14.31
	2	0.13009	-1437.07	2882.13	0
	3	0.13009	-1437.02	2884.04	1.913
	4	0.13003	-1436.19	2884.37	2.241
	5	0.13002	-1436.10	2886.20	4.070

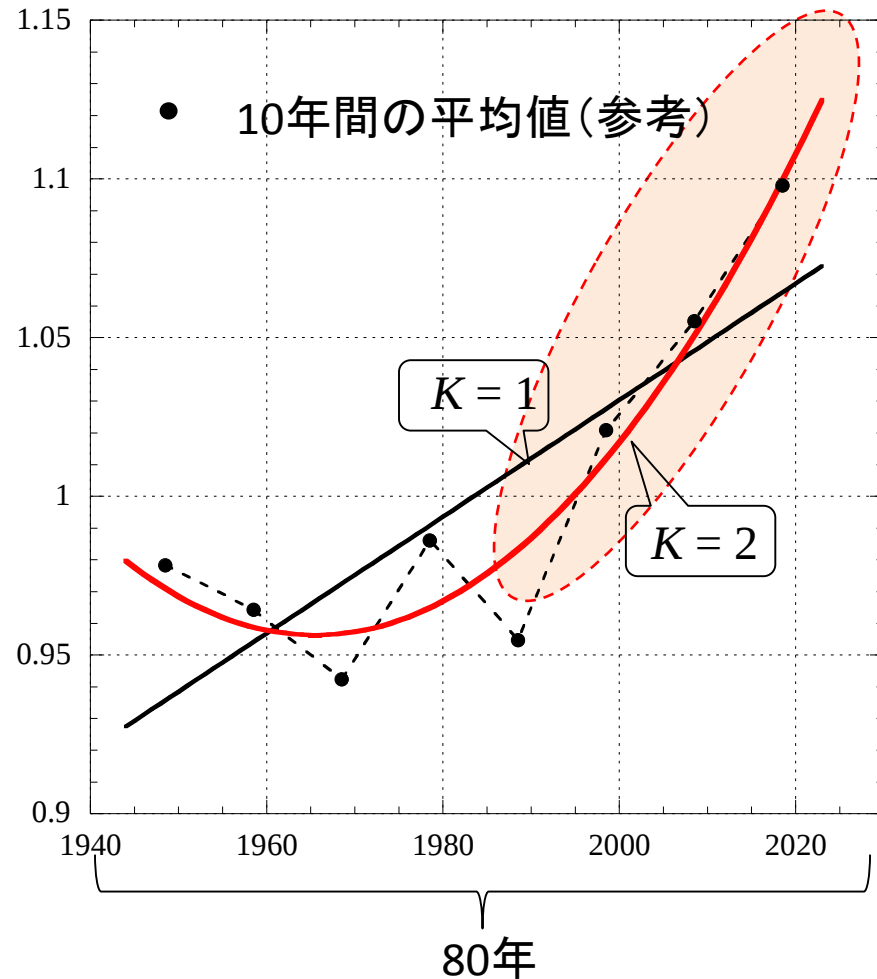
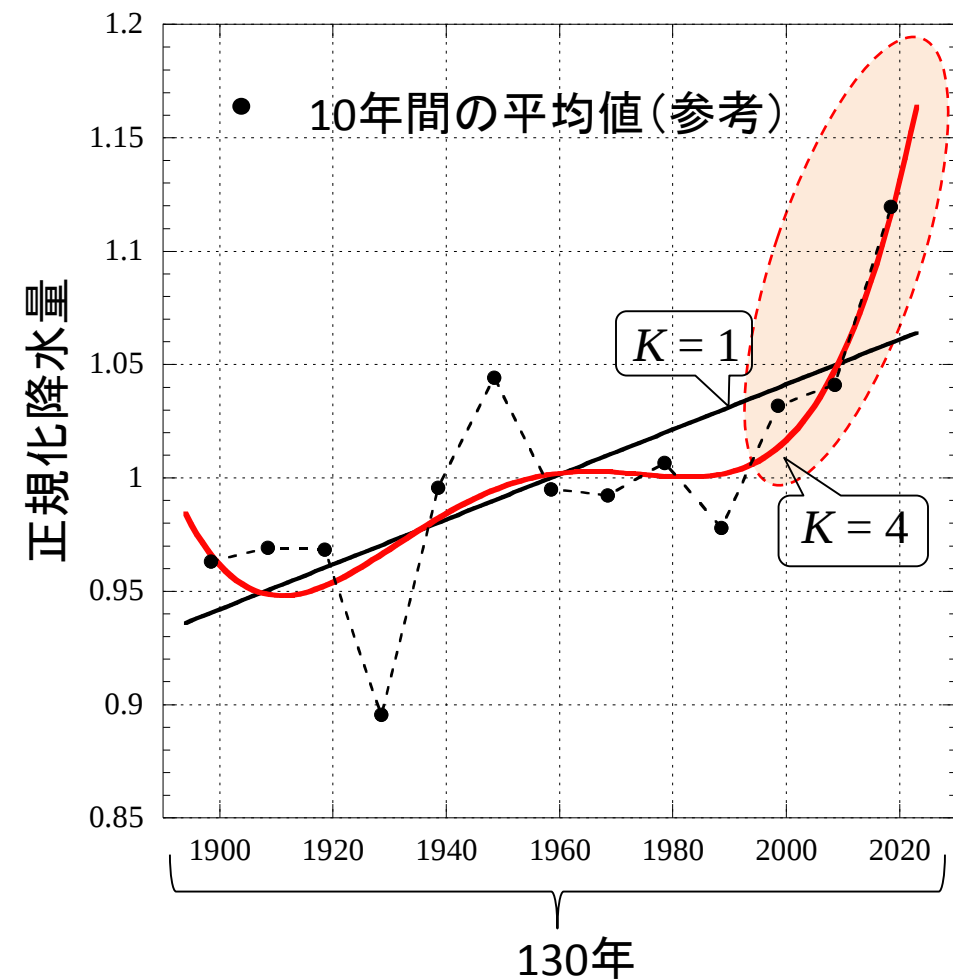
日降水量
45箇所
130年間の
データ

時間降水量
45箇所
80年間の
データ

年間最大雨量の長期変化傾向 (AICが選んだ最良次数)

【日降水量: 130年間】

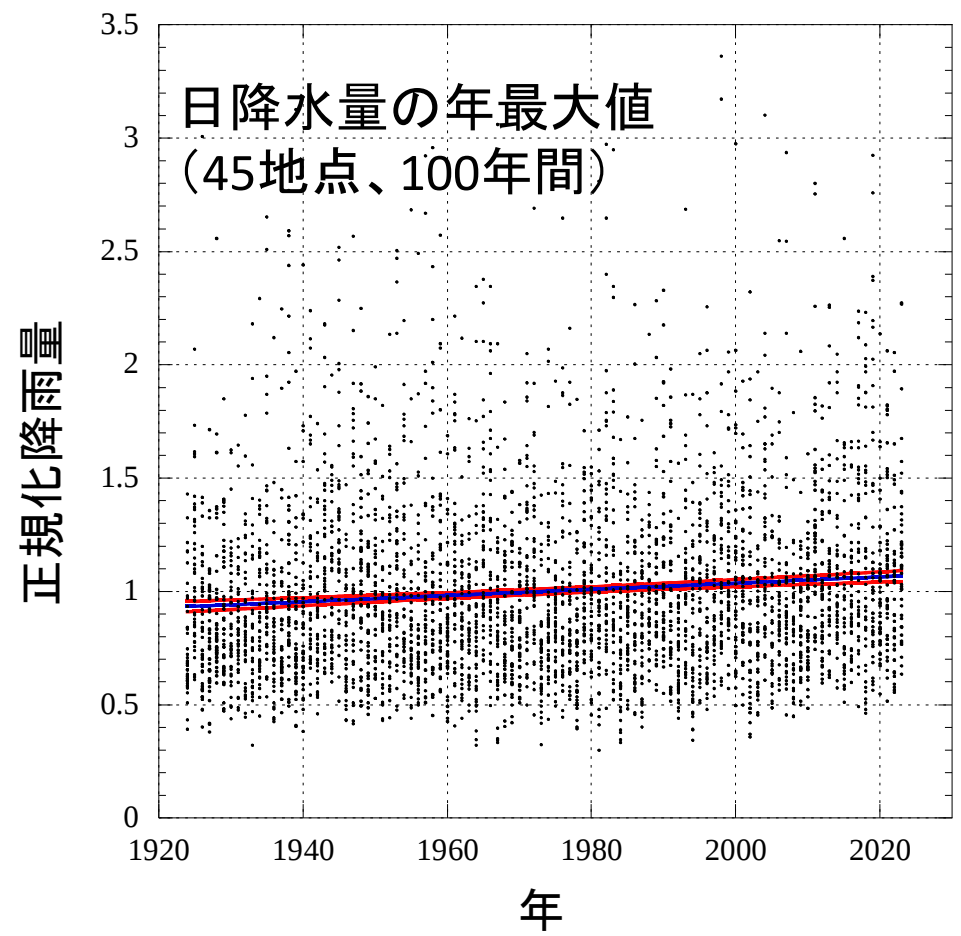
【時間降水量: 80年間】



直線回帰では100年間で10数%の増加であったが、
回帰曲線推定では、その増加は、1990年以降に顕著になっているように見える 62

以上の検討により、1980～1990年以降、雨の降り方は変わりつつある
(＝増加傾向にある)

とは言え



その変化の程度は、年毎のばら
つきの大きさに比べれば、十分小
さいので、
過去100年スケールで見れば、雨
の降り方は、定常状態(定常確率
過程)にあると言えるかもしれない

それなら、この状態がこのまま
1000年続くとしたら、1000年に一度
の大雨が予想できるであろうか？

極値統計

想定外の出来事を想定外としないために

例えば、100年に一度、1000年に一度の最悪事態
(ここでは豪雨)がどのくらいかを知りたい

最大値だけを集めたデータから、それを予想する

そのために、最大値の確率分布が分かれば良い

極値統計

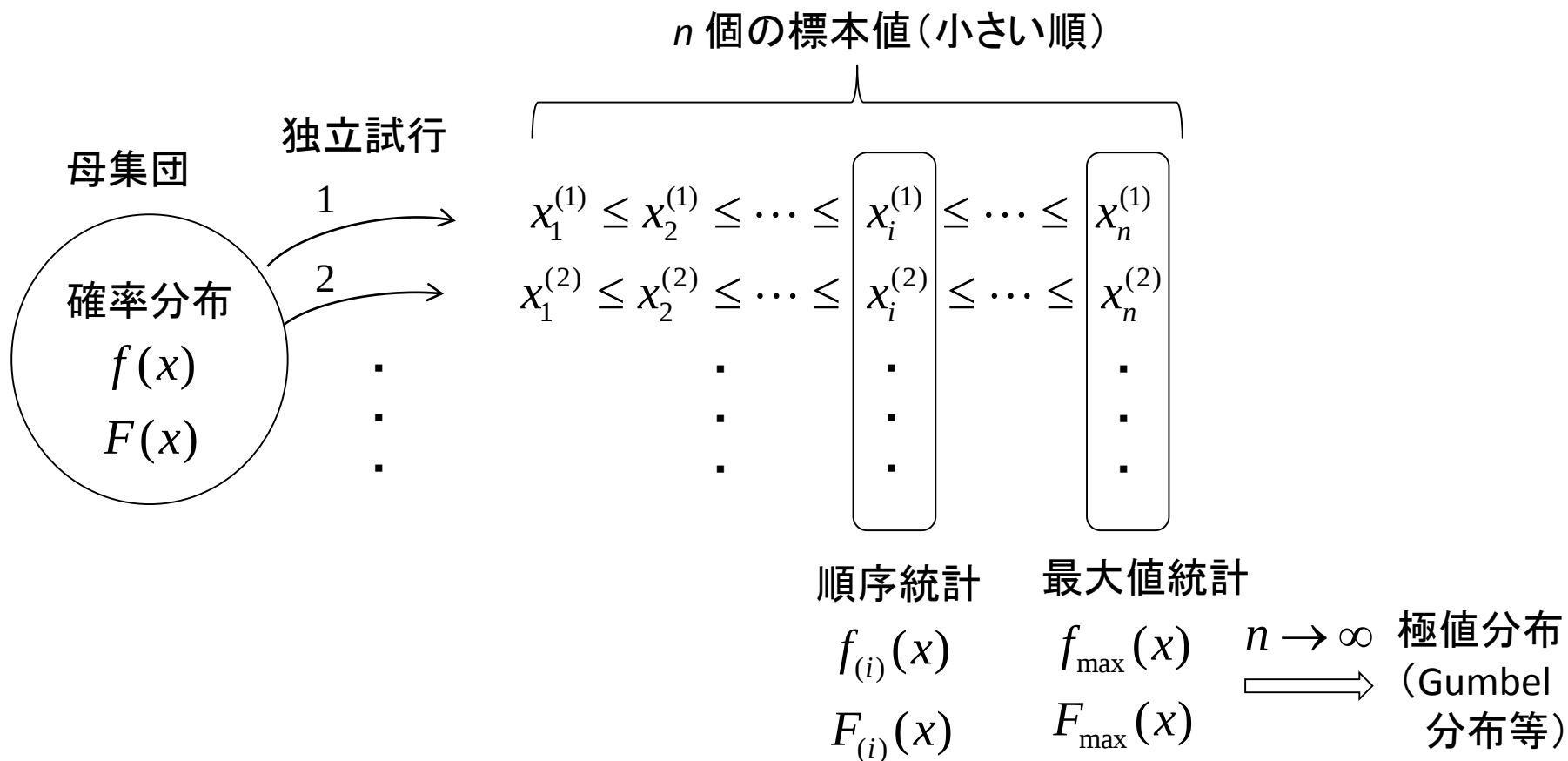
極値 $\equiv$ 最大値

たくさんのメンバーの中の最大値には、
元の分布(=メンバーの分布)によらない最大値特有の
確率分布がある

→ 極値分布

→ 最大値ばかりを集めたデータから、
非常にまれな出来事を予想する

順序統計・最大値統計・極値分布



順序 i 番目の分布と最大値の分布

母集団の確率分布: f : 確率密度関数、 F : 累積分布関数

n 個のサンプル値の小さいほうから i 番目の値の確率分布 (順序統計)

$$F_{(i)}(x) = \sum_{k=i}^n \binom{n}{k} F^k(x) \{1 - F(x)\}^{n-k}$$

$$f_{(i)}(x) = n \binom{n-1}{i-1} F^{i-1}(x) \{1 - F(x)\}^{n-i} f(x)$$

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!} \quad (0 \leq k \leq n)$$

最大値の分布 ($i = n$ より)

$$F_{\max}(x) = F_{(n)}(x) = F^n(x)$$

$$f_{\max}(x) = f_{(n)}(x) = nF^{n-1}(x)f(x)$$

元(母集団)の分布形が
与えられれば
具体的な形に記述できる

極値分布 (標本数が多いときの最大値に現れる分布)

最大値の累積確率分布

$$F_{\max}(x) = F_{(n)}(x) = F^n(x)$$

$n \rightarrow \infty$ の極限において、 F の形により、3つの分布系(極値分布)に収斂する

x の値が大きい部分(分布の裾)が指数関数的なものは、Gumbel分布に収斂する。電波伝搬に現れる大部分の分布(正規分布、指数分布、ガンマ分布、ワイブル分布、対数正規分布など)の極値分布はGumbel分布になる

Gumbel分布(グンベル分布)

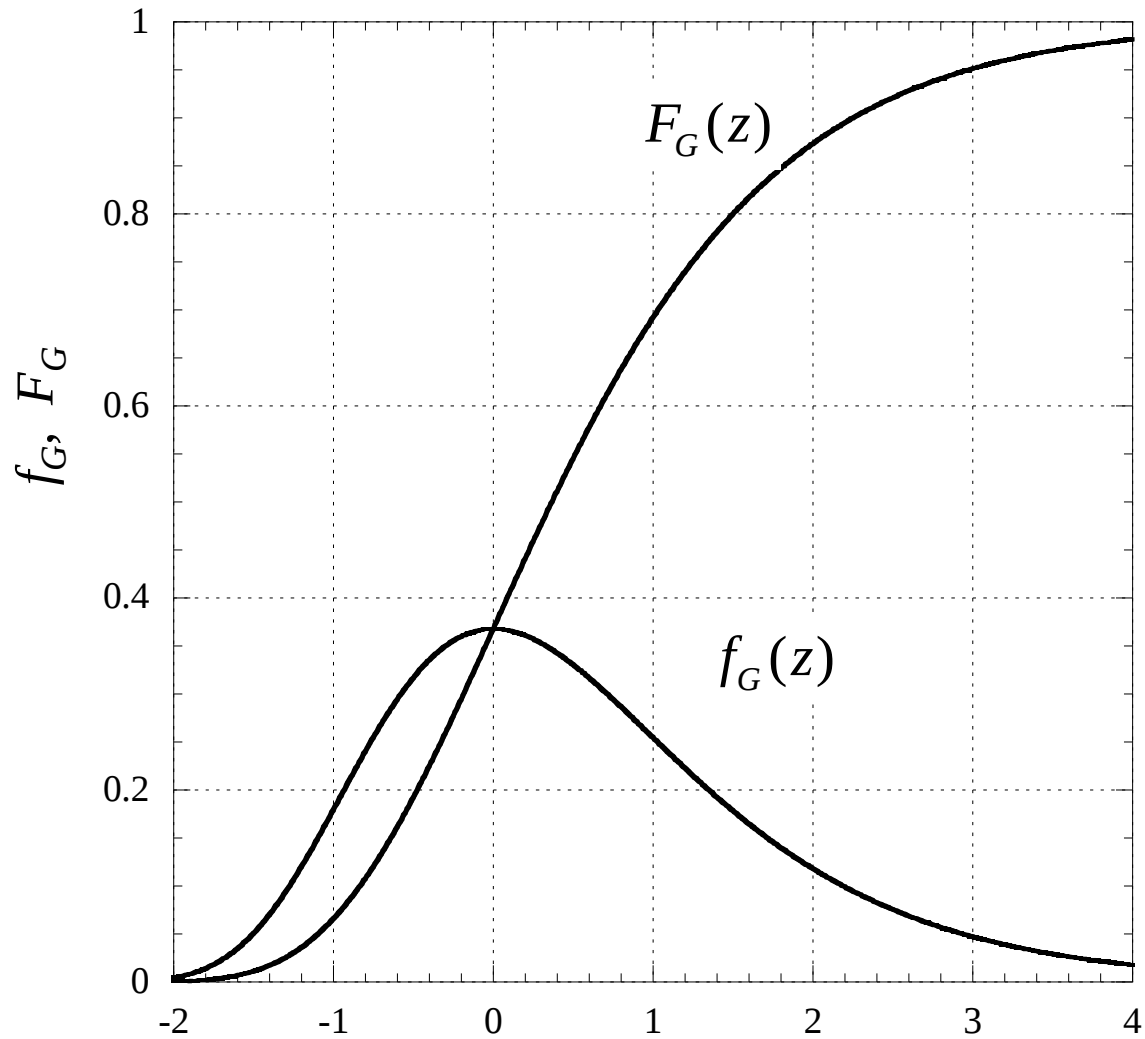
$$F_G(z) = \exp\{-\exp(-z)\}$$

$$f_G(z) = \exp\{-z - \exp(-z)\}$$

$$z = \frac{x_{\max} - b_n}{a_n}$$

分布のパラメータ a_n, b_n は
吸引係数と呼ばれる

極値分布: グンベル分布の形



$$z \{= (x_{\max} - b_n) / a_n ; n \rightarrow \infty\}$$

N年間に一度の最大値を求める

観測値

各地点のデータをN年ごとに分割してグループを作り、グループ毎の最大値を求めて平均値を出す

例えば、100年間のデータを持つ日降水量に対して、 $N=50$ の場合、地点ごとに2個、全部で90個の最大値が得られ、その平均値を求める。

極値分布からの推定

1) 全データに対する吸引係数 a_n , b_n を求める

(日降水量: 2点整合法; 時間降水量と10分間降水量: 平均値と分散から)

2) 平均値を求める

平均値(数値積分)

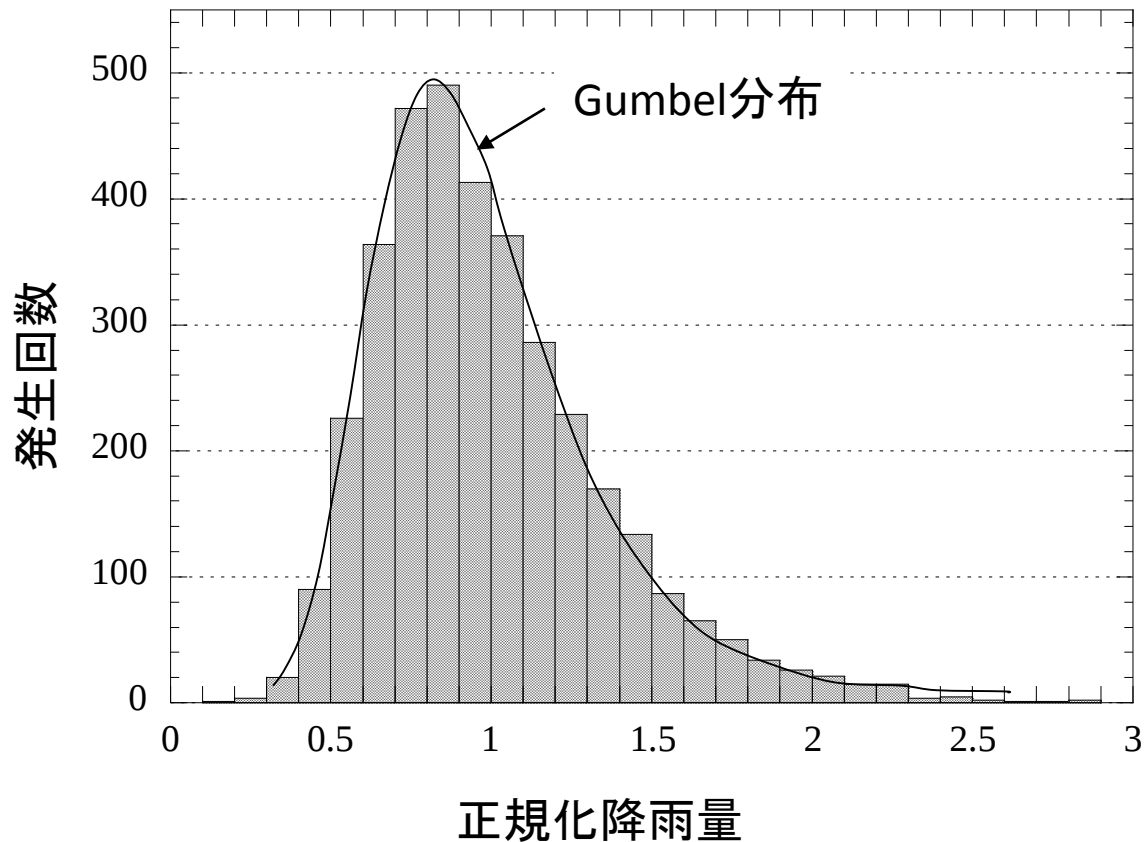
$$\langle x_{\max} \rangle = \int_{-\infty}^{\infty} x f_{\max}(x) dx \implies x_N = \frac{N}{a_n} \int_0^{\infty} x \{F_G(z(x))\}^{N-1} f_G(z(x)) dx$$

近似値(閉形式が得られる)

$$x_N \approx F_G^{-1}\left(\frac{N}{N+1}\right) = -a_n \log \left\{ -\log \left(\frac{N}{N+1} \right) \right\} + b_n$$

正規化降雨量年間最大値のヒストグラムと Gumbel分布でのフィッティング

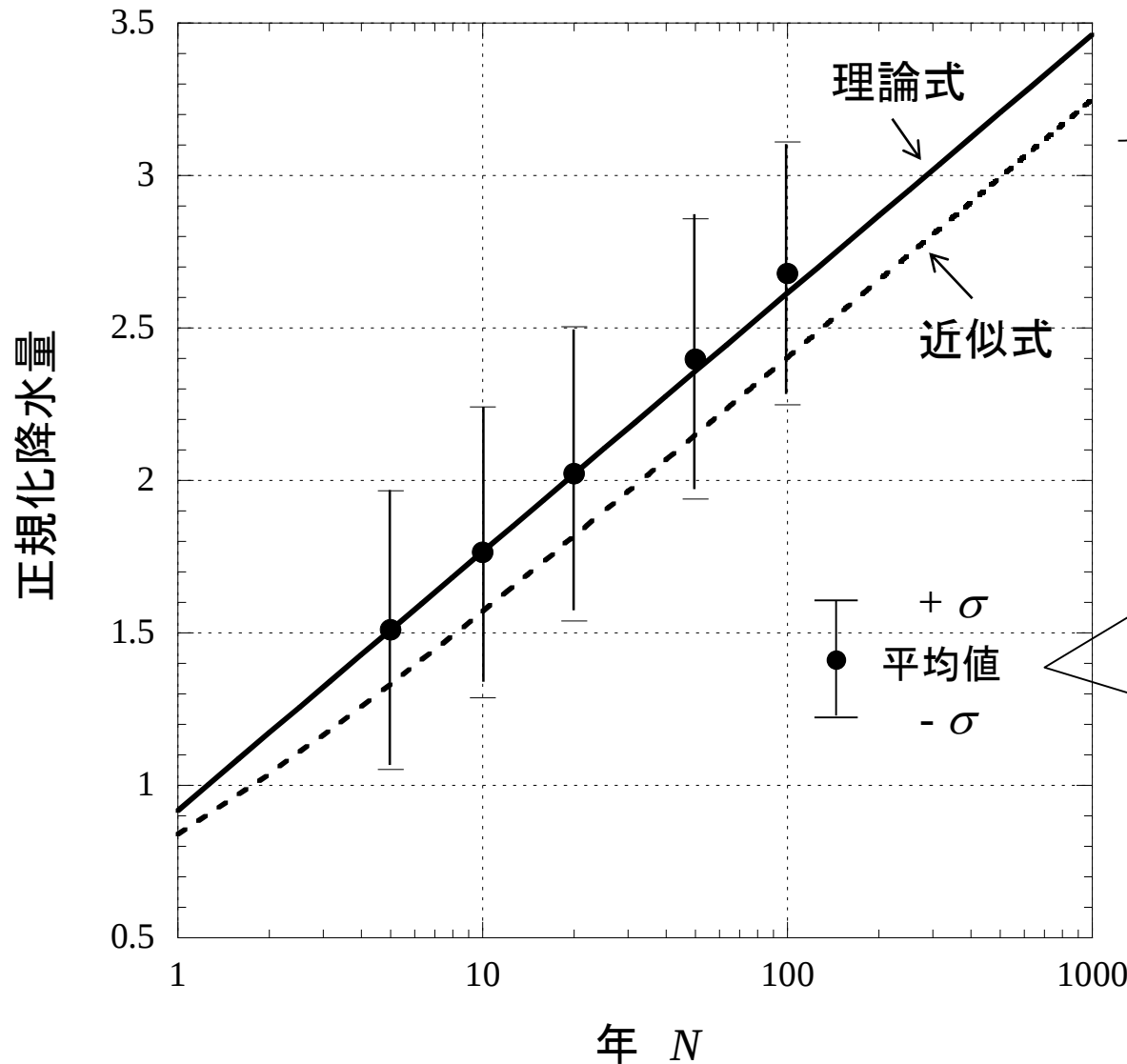
時間降水量の年間最大値(データ数: $80(\text{年}) \times 45(\text{地点}) = 3600$)



1時間降水量の年間最大値
であるので、 $n = 8,760$ 個の
中の最大値の集まり

年間最大値データの平均と
分散が得られれば、
Gumbel分布は定められる

正規化日降水量の N 年間最大値



正規化最大値データの
平均と分散を求め
Gumbel分布から
理論計算により推定

100年間のデータを N 年
ごとにグループ分けし、
その中の最大値(すな
わち最大値の中の最
大値)の平均値

この100年間に限ってみれば、どんな大雨も、一つの確率分布(=Gumbel分布)に従って起きている=想定外なことは起きていない

1000年に一度の最悪値が予測できるか？

① 手持ちのデータ(時間降水量の例で)

$$45\text{地点} \times 80\text{年間} = 3200(\text{データ数})$$

② 定常不規則過程を想定して、全地点データを一系列に並べて、3200年間のデータとみなす

$$1\text{地点(仮想)} \times 3200\text{年間(仮想)} = 3200(\text{仮想年数})$$

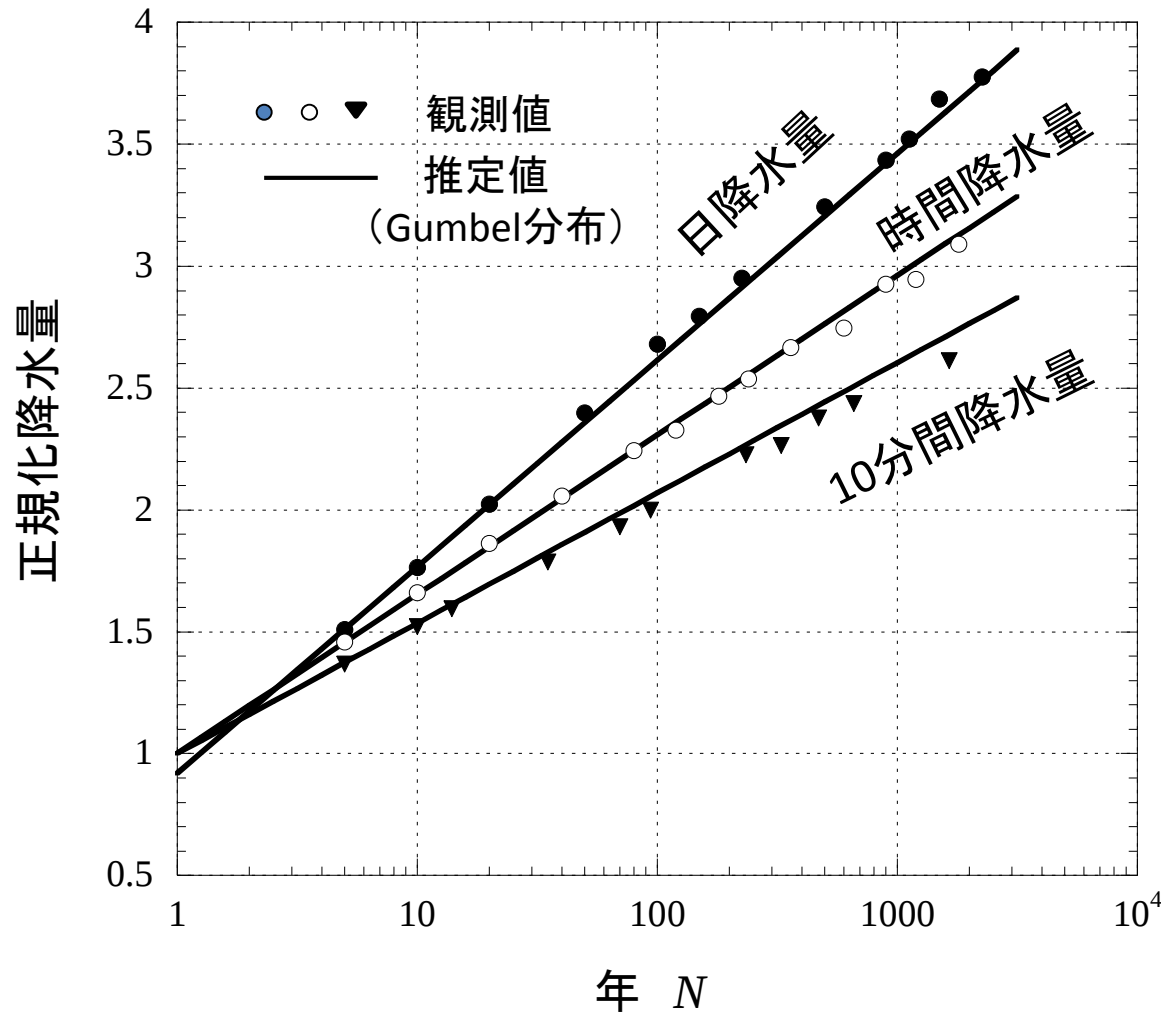
③ 理論的推定(極値統計)

①から平均と標準偏差を求める

→ 極値分布(Gumbel分布)から計算によって求める

理論推定の検証： ②と③の結果を比較する

多点正規化雨量をつなげて一点での仮想長期間データとして



日降水量: 4500年分
 時間降水量: 3600年分
 10分間降水量: 3290年分

Gumbel分布からの推定式

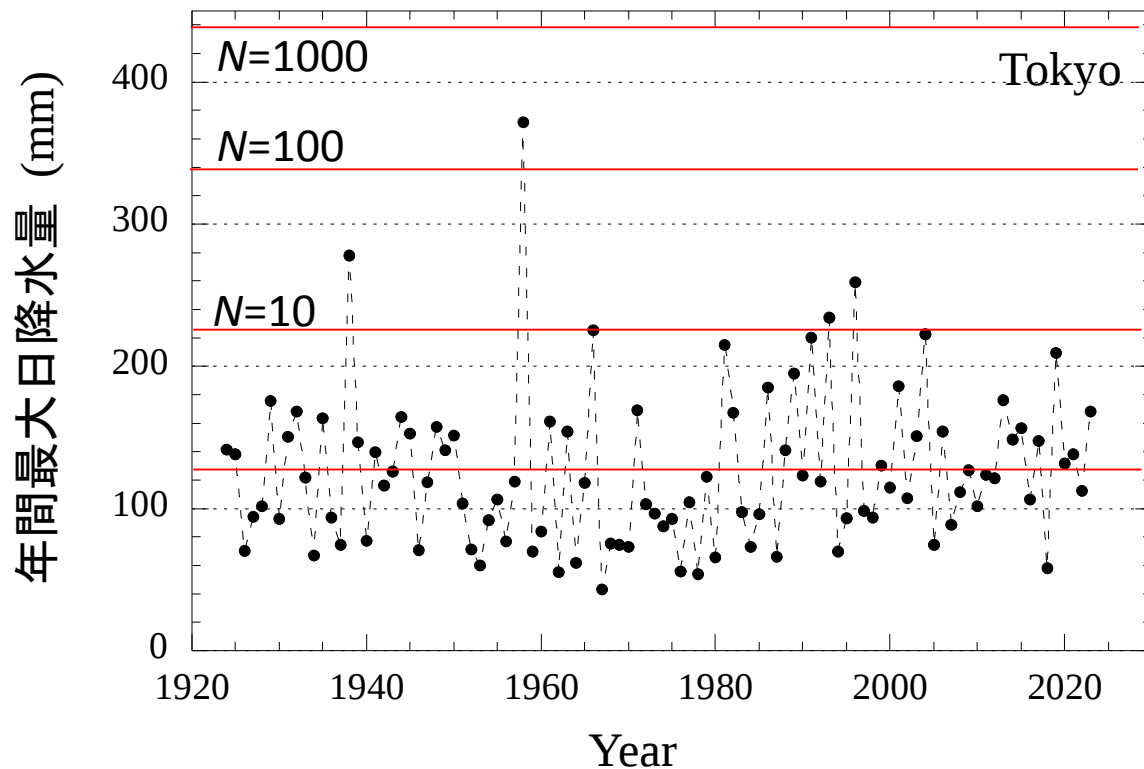
$$\langle R_{\max}(N) \rangle = (a \log_{10} N + 1) \langle R_{\max}(1) \rangle$$

$$a = \begin{cases} 0.83 & (1\text{-day rainfall}) \\ 0.66 & (1\text{-hour rainfall}) \\ 0.53 & (10\text{-minute rainfall}) \end{cases}$$

過去100年間を定常状態とみなして、これが1000年間続くとしたら、その間の
 N年間最悪値は、極値分布(Gumbel分布)で精度よく予測できる

東京における年間最大日降水量の100年間の推移と N 年間最大値の推定値

東京(千代田区)



100年間平均値
125.7 mm

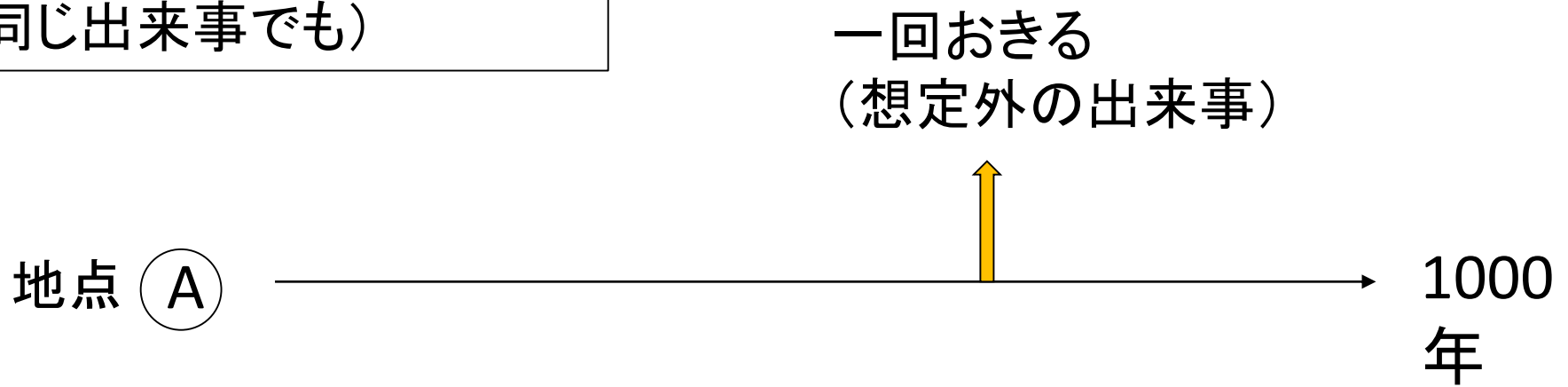
推定値

$N=10$	226 mm
$N=100$	339 mm
$N=1000$	440 mm

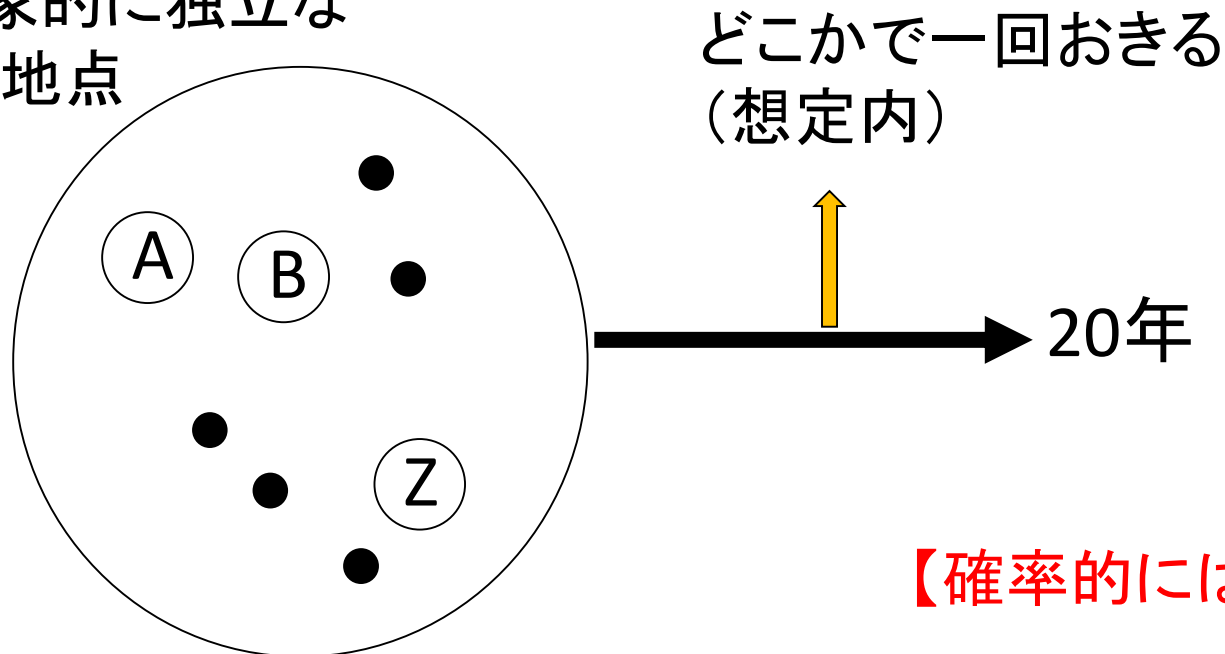
観測値

100年間最大	371.9 mm (1958)
2番目	287.3 mm (1938)

パラレルワールドの考え方
(同じ出来事でも)



気象的に独立な
50地点



【確率的には同じこと】

第2部：統計的推定のまとめ

三つの統計的推定法を取り上げ、無線通信にも関係ある雨の降り方の統計解析に応用した。日本各地(45地点)の日降水量と時間降水量の年間最大値の約100年分について解析した結果

- ① 地球温暖化の時代と言われるが、雨については、地点毎に100年スケールで変化を追ってみても、年毎のばらつきが大きく、変化傾向を読み取ることができない(回帰分析、区間推定)
- ② 日本各地の年間最大降水量をその地の平均値で正規化して、日本全国のデータをまとめて解析すると、100年間で十数%の増加傾向が見られ、95%の信頼区間推定においてもそこ傾向は有意であった(回帰分析、区間推定)
- ③ より詳しい経年変化を調べるために、多項式近似のモデルを用いて評価を行った結果、増加傾向は主に1980～1990年以降に顕著であることが明らかになった(多項式回帰分析、AIC)
- ④ 長期的変化が年毎の変動に比べて十分小さいので、100年単位で見たとき、一つの定常確率過程の出来事と捉えることができる。極値統計を使えば、最大値データから、10年・100年・1000年間の最悪値が予想でき、実測データも理論の予測どおりになっている。故に、近年、日本各地に起きている豪雨は、地球温暖化の影響があるかもしれないが、判断は慎重になったほうが良い(極値統計)

本日の講義に関連する参考資料情報

スライドを使った講義は、図がベースであり、分かった気分になるかもしれないが、浅い理解で終わってしまうことが多いし、復習しようとしてもスライド資料だけでは不十分である。さらに理解を深めるためには、文章や式でまとめられた専門書などで復習をしてほしい。

今回の講演は、以下の内容をベースにしています

電波伝搬モデルの基礎：確率分布と統計的推定

唐沢、ネット公開私製本

http://www.radio3.ee.uec.ac.jp/ronbun/TR-YK-078_Probability_Distributions.pdf

電波技術・無線通信に関する各種技術レポートを唐沢研究室ホームページから公開しています <http://www.radio3.ee.uec.ac.jp/>

本講演に関するご質問・ご意見は E-mail: karasawa@mail.uec.jp まで