

# R と Excel を用いた分布推定の実践例

近藤宏樹<sup>1</sup> 斎藤新悟<sup>2</sup>

<sup>1</sup> 日新火災海上保険株式会社

<sup>2</sup> 九州大学マス・フォア・インダストリ研究所

2011/11/09

独立同分布なデータから母集団の確率分布を推定する状況は多い。

## 分布推定の手順

- 分布形の候補を定める（例：正規分布，ガンマ分布）。
- 各分布形に対して，パラメータを推定する（例：最尤法，モーメント法）。
- フィットの良さを比較し，採用する確率分布を決定する。

## 本日の内容

- (1) 推定プロセスの R・Excel を用いた実践例。
- (2) 分布を推定していることに起因するリスク  
(パラメータリスク・モデルリスク) の評価。

## R とは

- オープンソースでフリーな統計解析ソフト.
- 様々な専門家により多くの統計的手法が実装されており, パッケージとして入手可能.
- グラフィック機能に優れている.

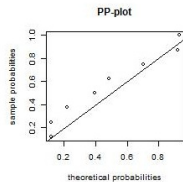
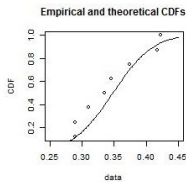
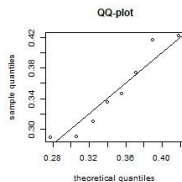
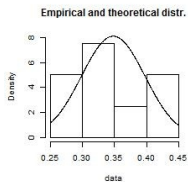
## R による分布推定の例

```
> library(fitdistrplus)
> x <- c(0.335,0.417,0.374,0.290,0.310,0.346,0.422,0.289)
> fitdist(x,"norm","mle")
Fitting of the distribution ' norm ' by maximum likelihood
Parameters:
      estimate Std. Error
mean 0.34787500 0.01739921
sd    0.04921239 0.01228025
```

## R を用いる利点

- 推定手順はすでにプログラムされている。
- フィットの良さを表す統計量の計算や図の表示も簡単にできる。

```
> fit <- fitdist(x,"norm","mle")
> gofstat(fit)
Kolmogorov-Smirnov statistic: 0.1699337
Cramer-von Mises statistic: 0.04572002
Anderson-Darling statistic: 0.3562003
> plot(fit)
```



## Excel の利用：出力結果の表示

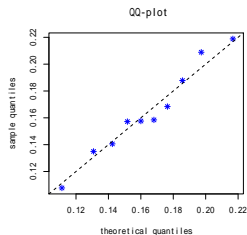
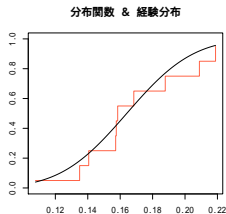
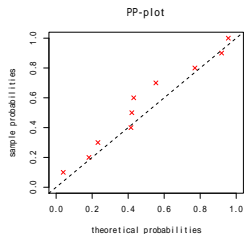
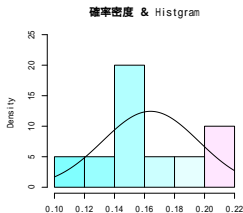
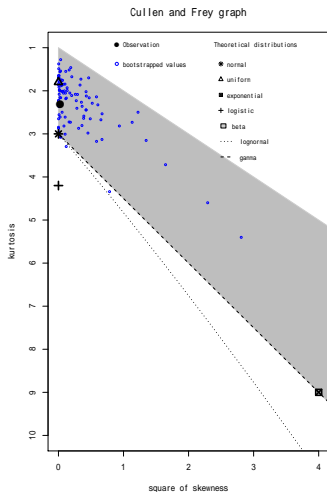
R の計算結果を視覚化するには Excel の表を用いると便利。

種目	正規分布 (モーメント法)						対数正規分布 (モーメント法)						分布選択		
	パラメータ		p値	AD検定		KS検定	パラメータ		p値	AD検定		KS検定	各統計量による選択		
	mean	sd		統計量	検定		統計量	meanlog		sdlog	統計量		検定	統計量	p値
種目1	0.52	0.01	0.3948	0.343	採択	0.158	-0.65	0.03	0.3883	0.328	採択	0.153	正規分布	対数正規分布	対数正規分布
種目2	0.33	0.06	0.4006	0.237	採択	0.141	-1.13	0.18	0.5263	0.196	採択	0.123	対数正規分布	対数正規分布	対数正規分布
種目3	0.28	0.04	0.2511	0.333	採択	0.191	-1.27	0.13	0.2227	0.268	採択	0.165	正規分布	対数正規分布	対数正規分布
種目4	0.16	0.03	0.0000	0.261	採択	0.169	-1.83	0.20	0.0000	0.236	採択	0.154	対数正規分布	対数正規分布	対数正規分布
種目5	0.41	0.07	0.0318	0.591	採択	0.266	-0.91	0.17	0.0688	0.537	採択	0.246	対数正規分布	対数正規分布	対数正規分布
種目6	0.29	0.05	0.2013	0.393	採択	0.194	-1.24	0.18	0.2029	0.368	採択	0.185	対数正規分布	対数正規分布	対数正規分布
種目7	0.21	0.06	0.0647	0.504	採択	0.194	-1.60	0.25	0.1497	0.294	採択	0.148	対数正規分布	対数正規分布	対数正規分布
種目8	0.51	0.05	0.1163	0.603	採択	0.221	-0.69	0.10	0.0988	0.712	採択	0.243	正規分布	正規分布	正規分布
種目9	0.48	0.08	0.4014	0.649	採択	0.271	-0.74	0.14	0.5492	0.507	採択	0.248	対数正規分布	対数正規分布	対数正規分布
種目10	0.37	0.07	非計算	0.218	採択	0.165	-1.02	0.18	非計算	0.173	採択	0.138	判定不能	対数正規分布	対数正規分布
種目11	0.12	0.02	0.3863	0.271	採択	0.149	-2.13	0.14	0.4858	0.183	採択	0.123	対数正規分布	対数正規分布	対数正規分布
種目12	0.15	0.03	0.6343	0.262	採択	0.143	-1.94	0.19	0.7452	0.261	採択	0.171	対数正規分布	対数正規分布	正規分布
種目13	0.22	0.13	0.0095	0.287	採択	0.178	-1.77	0.77	0.0008	0.375	採択	0.210	正規分布	正規分布	正規分布
種目14	0.51	0.06	0.5161	0.431	採択	0.237	-0.68	0.12	0.6331	0.413	採択	0.233	対数正規分布	対数正規分布	対数正規分布
種目15	0.34	0.09	0.0349	0.317	採択	0.180	-1.12	0.31	0.0157	0.453	採択	0.239	正規分布	正規分布	正規分布
種目16	0.69	0.30	0.3109	0.542	採択	0.212	-0.46	0.42	0.6596	0.223	採択	0.135	対数正規分布	対数正規分布	対数正規分布
種目17	0.09	0.03	0.1686	0.323	採択	0.159	-2.44	0.31	0.0885	0.507	採択	0.221	正規分布	正規分布	正規分布

フィットの良さを表す統計量は複数あり、  
統計量によって最良と判断される分布が異なることもある。  
→ 統計量以外の判断材料：図の利用。

## フィットの良さを表す図

### 種目4 正規分布 / モーメント法



## Excel の利用 : R コードの生成

- Excel の数式によって R コードを生成する。  
→ 生成されたコードをコピーして R で実行。
- プルダウンメニューなどを利用した直感的な操作も可能。

種目	種目4	▼
確率分布	正規分布	▼
推定方法	モーメント法	▼
作業ディレクトリ	D:\home	

青地の部分 は、  
入力に応じて変化する。

```
library(fitdistrplus)
setwd("D:/home") # set working directory
source("redef_descdist.r")
pdf("種目4norm_mme.pdf", family="Japan1", paper="a4r", width=11)
data <- read.csv("data.csv"); item_vec <- names(data)
i <- 4
item <- item_vec[i]
eval(parse(text=paste("x <- data$", item, "[!is.na(data$", item, ")"]", sep="")))
y <- sort(x)
k.n <- (order(y)-0.5)/length(y)
obsp <- ecdf(y)(y)
par(oma=c(0, 0.5, 4, 0.5))
layout(matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 2, 3, 2, 3, 4, 5, 4, 5), 2, 10, byrow = F))
#Oulien&Fray
descdist(x, boot=100)
#PDF
yl <- max(density(x)$y)
hist(x, breaks = "Sturges", col=cm.colors(10), prob=T, xlab="", main="確率密度 & Histogram", ylim=c(0, yl*2))
#CDF
curve(pnorm(x, 0.1640474742, 0.0320630550964756), add=TRUE, col=1, lty=1)
#P-P & Q-Q
theo <- pnorm(y, 0.1640474742, 0.0320630550964756)
plot(theo, obsp, main = "PP-plot", xlab = "theoretical probabilities", ylab = "sample probabilities",
xlim=c(0, 1), ylim=c(0, 1), cex = 1, col = 2, pch=4)
plot(theo, y, main = "QQ-plot", xlab = "theoretical quantiles", ylab = "sample quantiles",
xlim=c(min(theo, y), max(theo, y)), ylim=c(min(theo, y), max(theo, y)), cex = 1, col = 4, pch=8)
abline(0, 1, lty=2)
mtext("種目4 \n 正規分布/モーメント法", outer=T, side=3, cex=1.5)
dev.off()
```

## 分布推定に伴うリスク

データから推定した分布に従って将来予測（例：保険金の予測）を行う。  
→ **推定分布と真の分布との差**に起因する誤差が生じる。

- **パラメータリスク**  
パラメータの推定誤差に起因するリスク。
- **モデルリスク**  
モデル（確率分布の種類）の選択誤りに起因するリスク。

これらのリスクを計量する方法を考える。



## 状況設定

- データは**正規分布**または**対数正規分布**に従うと仮定.
- パラメータリスク・モデルリスクを織り込んだリスク指標（例：VaR）を計算したい.

## Bayes 推定の利用

- どちらの分布に従うか,
- パラメータがどの値を取るか

が確率的に変動すると考える.

**Bayes の定理**によって、分布・パラメータの確率変動がデータを加味して計算できる.

## Rによる実装

- データから分布・パラメータの確率変動を計算。
- 分布・パラメータの確率変動を含んだリスク指標を計算。  
→ Rによるモンテカルロ・シミュレーションを用いた。

回数	分布	期待値	標準偏差	$x$
1	正規	0.52	0.13	0.44
2	正規	0.39	0.15	0.21
3	対数正規	0.38	0.08	0.45
⋮	⋮	⋮	⋮	⋮

## 数値例

データ（ある保険種目の損害率を想定）：

33.5%, 41.7%, 37.4%, 29.0%, 31.0%, 34.6%, 42.2%, 28.9%, 23.0%, 27.2%.

算出結果：

	(a)	(b)	(c)	(d)	(e)
期待値	32.8%	32.9%	32.8%	33.0%	32.9%
99%VaR	46.7%	49.5%	49.0%	53.1%	50.0%

- (a) 正規分布を仮定，パラメータを固定（最尤法による推定値）。
- (b) 対数正規分布を仮定，パラメータを固定（最尤法による推定値）。
- (c) 正規分布を仮定，パラメータリスクを考慮。
- (d) 対数正規分布を仮定，パラメータリスクを考慮。
- (e) パラメータリスクおよびモデルリスクを考慮。

## R と Excel を用いた分布推定

- 分布推定にあたっては、R の機能が有効に活用できる。
- Excel を併用することで GUI 機能を補う方法を提示した。

## パラメータリスク・モデルリスク

- 特定の状況下でパラメータリスク・モデルリスクを計量した。
- しかし、実務に応用するためには課題も多い。  
(異なる確率分布を使用する場合はモデルリスクの計算が困難.)