

はじめての統計データ分析

豊田秀樹

2016 年（朝倉書店）

1 データの整理とベイズの定理

1.4 ベイズの定理

1.4.2 事前分布

本書の範囲では、複数の母数は互いに独立であることを仮定する。例えば正規分布における同時事前分布は、

$$f(\boldsymbol{\theta}) = f(\mu, \sigma) = f(\mu)f(\sigma) \quad (1.33)$$

と書ける。

無情報の事前分布として一様分布を利用する場合、区間を無限大にすると確率密度は 0 になり、変則分布（変則事前分布）と呼ばれるものになる。しかし、変則事前分布を用いても、事後分布は発散することなくベイズの定理を利用できることが知られている¹。

2 MCMC と正規分布の推測

2.1 マルコフ連鎖モンテカルロ法

2.1.2 事前分布の設定

- Gibbs sampling（GS 法）：特定の母数には特定の事前分布の型を指定する必要がある。
- ハミルトニアンモンテカルロ（HMC）法：分析者が自分の意思で自由な事前分布を選ぶことができる。

2.1.5 乱数列の数値的評価

有効標本数 n_{eff} 発生させた乱数が理想的な乱数のいくつ分に相当するか²。標準誤差を算出する際に用いる。

収束判定指標 \hat{R} チェイン間と内の散らばりを比較することで、事後分布から乱数を発生させているかを判定する指標。

Table 2.1: 「牛井問題」の乱数列の評価

	n_{eff}	\hat{R}
μ	32600	1.00
σ	29714	1.00

¹詳細については記述なし。要確認。

²MCMC で発生させる乱数は系列相関を有する。

2.2 事後分布

2.2.2 事後分散・事後標準偏差

事後標準偏差 θ の標準偏差。母数 θ が $\hat{\theta}_{eqp}$ の周辺でどれほど散らばっているか。 n に依存して値が決まる。

2.2.3 事後期待値の標準誤差

標準誤差 $\hat{\theta}_{eqp}$ の標準偏差。乱数の数 T に依存して値が決まる。

$$\hat{\theta}_{eqp} \sim N(\theta_{eqp}, \sigma_{\theta} / \sqrt{n_{eff}}) = N(\theta_{eqp}, \sigma_{\hat{\theta}_{eqp}}) \quad (2.9)$$

2.4 母数と予測分布に関するベイズ的推測

実質科学的知見 統計学とは全く関係のない知見（固有技術、ドメイン知識）

何らかの基準点を設定する場合、実質科学的知見からの要請にもとづく場合が多い。

2.5 生成量

生成量 MCMC 法による標本 $\theta^{(t)}$ の関数 $g(\theta^{(t)})$ 。

2.5.1 分散

$$g(\sigma^{(t)}) = \sigma^{(t)2} \quad (2.20)$$

メモ

生成量としてではなく、最初から分散を母数としてはダメなのか？（本文では、正規分布の母数にも分散ではなく標準偏差が用いられている。）

メモここまで

2.5.2 変動係数

測定値が比尺度である場合の散布度の指標³。

$$g(\mu^{(t)}, \sigma^{(t)}) = \frac{\sigma^{(t)}}{\mu^{(t)}} \quad (2.24)$$

メモ

このように複数の母数が組み合わさった生成量については、 $\mu^{(t)}$ と $\sigma^{(t)}$ が同時分布からの標本 $\theta^{(t)}$ として得られている場合にのみ意味がある。例えば、Gibbs sampling のように母数ごとの周辺分布からの標本しか得られない場合は、このような生成量を構成しても無意味である（と思う）。

メモここまで

2.5.3 効果量

$$\delta_c^{(t)} = g(\mu^{(t)}, \sigma^{(t)}) = \frac{\mu^{(t)} - c}{\sigma^{(t)}} \quad (2.26)$$

ちなみに、定数 c を変化させた場合の効果量の事後分布は単純に平行移動したものにならないので注意。

³原点が固定されていないと、意味をなさない。

2.5.4 分位点・%点

モデル分布が正規分布のように既知である場合、予測値 x^* の%点の事後分布を

$$g(\mu^{(t)}, \sigma^{(t)}) = \mu^{(t)} + z \times \sigma^{(t)} \quad (2.27)$$

のように求めることができる⁴。

2.5.5 予測分布の特定区間の確率

将来のデータが区間 $[a, b]$ に観察される予測確率

$$p_{ab} = F(b|\mathbf{x}) - F(a|\mathbf{x}) \quad (2.28)$$

の事後分布を、生成量

$$g(\mu^{(t)}, \sigma^{(t)}) = F(b|\mu^{(t)}, \sigma^{(t)}) - F(a|\mu^{(t)}, \sigma^{(t)}) \quad (2.29)$$

で求める。

事後予測分布と条件付き予測分布

- 事後予測分布：%点は推定値（1 つだけ）としてしか評価できない
- 条件付き予測分布：%点は分布として評価できる

メモ

一般に、事後予測分布は既知の分布になるとは限らない。一方、条件付き予測分布であれば、母数の候補値を所与とした既知のモデル分布になる。

メモここまで

2.5.6 基準点との比

$$g(x^{*(t)}) = x^{*(t)} / c \quad (2.30)$$

2.8 章末問題

stan のコード

```
data {  
  int<lower=0> N;  
  vector<lower=0>[1] x[N];  
}  
  
parameters {  
  real mu;  
  real<lower=0> sigma;  
}  
  
model{  
  for(i in 1:N){  
    x[i] ~ normal(mu,sigma);  
  }  
}
```

⁴ 予測分布の%点の事後分布なので、メタ分布。気象庁が降水確率に確信度 ABC を付帯させているという例が提示されている。

```

generated quantities{
  real<lower=0> sigmasq; //分散
  real x_star; //予測値
  real cv; //変動係数
  real delta_35; //効果量 (基準値 c=35)
  real<lower=0,upper=1> px_under35_1; //35 点未満の値が観測される事後予測確率
  real rto_35; //基準点 (c=35) との比
  real<lower=0,upper=1> pmu_under35; //母平均が 35 点未満の確率
  real<lower=0,upper=1> px_under35_2; //35 点未満の値が観測される事後予測確率
  real<lower=0,upper=1> pdelta_under1; //効果量が 1.0 未満である確率
  real<lower=0,upper=1> px_under35_over90; //35 点未満の値が観測される確率が 90%以上の確率
  sigmasq = pow(sigma,2);
  x_star = normal_rng(mu,sigma);
  cv = sigma/mu;
  delta_35 = (mu-35)/sigma;
  px_under35_1 = normal_cdf(35,mu,sigma);
  rto_35 = x_star/35;
  pmu_under35 = mu<35 ? 1:0;
  px_under35_2 = x_star<35 ? 1:0;
  pdelta_under1 = delta_35<1.0 ? 1:0;
  px_under35_over90 = px_under35_1>0.9 ? 1:0;
}

```

R のコード

```

library("rstan")
# 第 2 章
# 章末問題
N <- 100
x <- structure(.Data=c(
  36,38,51,40,41,52,43,31,35,37,49,43,43,41,36,53,43,26,45,37,
  33,38,33,35,36,28,46,41,32,49,43,38,46,46,46,45,44,40,38,37,
  35,39,31,55,48,32,37,37,45,39,42,40,40,50,38,51,29,44,41,42,
  43,36,38,33,32,42,43,40,46,54,37,24,47,35,35,47,38,31,41,39,
  40,43,37,45,38,42,48,43,38,48,47,44,42,36,50,36,55,51,38,33),.Dim=c(100,1))
data01 <- list(N=N,x=x)
scr <- "toyoda2016-exer02.stan"
par <- c("mu","sigma","sigmasq","x_star","cv","delta_35","px_under35_1","rto_35",
  "pmu_under35","px_under35_2","pdelta_under1","px_under35_over90")
war <- 5000
ite <- 105000
see <- 1234
dig <- 3
cha <- 4
fit01 <- stan(file=scr,model_name=scr,data=data01,pars=par,verbose=F,seed=see,
  chains=cha,warmup=war,iter=ite)
print(fit01,pars=par,digits_summary=dig)
plot(fit01,pars=par)
traceplot(fit01,inc_warmup=T,pars=par)

```

実行結果

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu	40.643	0.004	0.650	39.368	40.206	40.642	41.083	41.919	34081	1
sigma	6.509	0.003	0.467	5.675	6.185	6.480	6.804	7.501	31680	1
sigmasq	42.584	0.035	6.158	32.210	38.252	41.993	46.289	56.270	31548	1
x_star	40.608	0.033	6.552	27.781	36.180	40.581	45.042	53.493	40000	1

cv	0.160	0.000	0.012	0.139	0.152	0.159	0.168	0.185	32226	1
delta_35	0.871	0.001	0.117	0.644	0.793	0.871	0.950	1.101	34035	1
px_under35_1	0.193	0.000	0.032	0.136	0.171	0.192	0.214	0.260	33600	1
rto_35	1.160	0.001	0.187	0.794	1.034	1.159	1.287	1.528	40000	1
pmu_under35	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	40000	NaN
px_under35_2	0.195	0.002	0.396	0.000	0.000	0.000	0.000	1.000	40000	1
pdelta_under1	0.864	0.002	0.343	0.000	1.000	1.000	1.000	1.000	27626	1
px_under35_over90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	40000	NaN

3 独立した2群の差の推測

3.4 非重複度

非重複度 (Cohen の U_3) 第1群⁵の平均値 μ_1 は、第2群では何%点に相当するか。

$$U_3 = F(\mu_1 | \mu_2, \sigma) \quad (3.17)$$

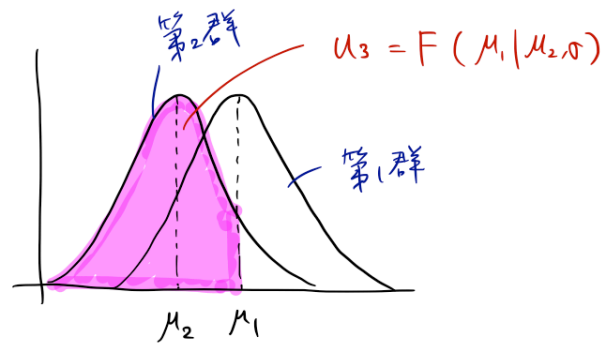


Figure 3.1: 非重複度

実験群と統制群で非重複度が 80% であれば、その処置によって平均的な生徒が全体の 30%(= 80% - 50%) を追い抜くと解釈できる⁶。

非重複度がある基準確率 c よりも大きい確率は、生成量

$$u_{U_3 > c}^{(t)} = g(\mu_1^{(t)}, \mu_2^{(2)}, \sigma^{(t)}) = \begin{cases} 1 & U_3^{(t)} > c \\ 0 & U_3^{(t)} \leq c \end{cases} \quad (3.21)$$

で計算することができる。

3.5 優越率

優越率 (probability of dominance) 無作為に選んだ一方の群の測定値が、無作為に選んだ他方の群の測定値を上回る確率。

$$x_1^* - x_2^* \sim N(\mu_1 - \mu_2, \sqrt{2}\sigma) \quad (3.22)$$

から、

$$\pi_d = p(x_1^* - x_2^* > 0) = p\left(z < \frac{\delta}{\sqrt{2}}\right) \quad (3.23)$$

と導かれる⁷。

⁵ 平均値の大きい方を第1群とする。

⁶ 正規分布の場合。

⁷ $z \sim N(0, 1)$

方法 1 生成量

$$\pi_d^{(t)} = g((\mu_1^{(t)}, \mu_2^{(2)}, \sigma^{(t)}) = F\left(\frac{\delta^{(t)}}{\sqrt{2}} \middle| \mu = 0, \sigma = 1\right) \quad (3.24)$$

を利用して、優越率の事後分布を求める⁸。

この生成量 $\pi_d^{(t)}$ を利用すると、生成量が基準確率 c よりも大きい確率を、生成量

$$u_{\pi_d > c}^{(t)} = g(\mu_1^{(t)}, \mu_2^{(2)}, \sigma^{(t)}) = \begin{cases} 1 & \pi_d^{(t)} > c \\ 0 & \pi_d^{(t)} \leq c \end{cases} \quad (3.25)$$

の EAP により評価することもできる。

方法 2 事後予測分布を利用し、生成量

$$u_{x_1^* - x_2^* > 0}^{(t)} = g(\mu_1^{(t)}, \mu_2^{(2)}, \sigma^{(t)}) = \begin{cases} 1 & x_1^* - x_2^* > 0 \\ 0 & x_1^* - x_2^* \leq 0 \end{cases} \quad (3.26)$$

の EAP を評価する⁹。

3.6 閾上率

閾上率 (probability beyond threshold) 無作為に選んだ一方の群の測定値と他方の群の測定値の差が閾値 c を上回る確率。

$$\pi_c = p(x_1^* - x_2^* > c) = p\left(z < \frac{\mu_1 - \mu_2 - c}{\sqrt{2}\sigma}\right) \quad (3.27)$$

優越率と同様、事後分布を直接評価、もしくは事後予測分布を通じて評価することができる。

3.8 標準偏差が異なるモデル

2 群の標準偏差を等しいという制約をおいた場合とおかなかった場合とでは、(いずれか片方の標準偏差を分母とした場合の) 標準化された平均値差 (効果量) と非重複度の値は食い違いが生じる。一方、優越率および閾上率にはあまり影響を及ぼさない。

一般に、当該制約が適切である場合は、「制約のあるモデル」の方が、母数の事後標準偏差は小さくなる。

WAIC (Widely Applicable Information Criterion, or Watanabe Akaike Information Criterion)

MCMC と相性の良い情報量規準。将来のデータ \mathbf{x}^* の予測の程度の指標で、WAIC の小さなモデルを「良いモデル」とする¹⁰。

3.9 章末問題

メモ

95%片側確信区間を求めるために、90%両側確信区間の下側、上側を求めたい。stan では上下 2.5%、25%、50%点のみデフォルトで報告される。ここに 5%を追加するにはどうしたらよいか？

メモここまで

⁸平均、SD、確信区間が求まる。

⁹点推定値しか得られない。

¹⁰Watanabe (2009; 2010)

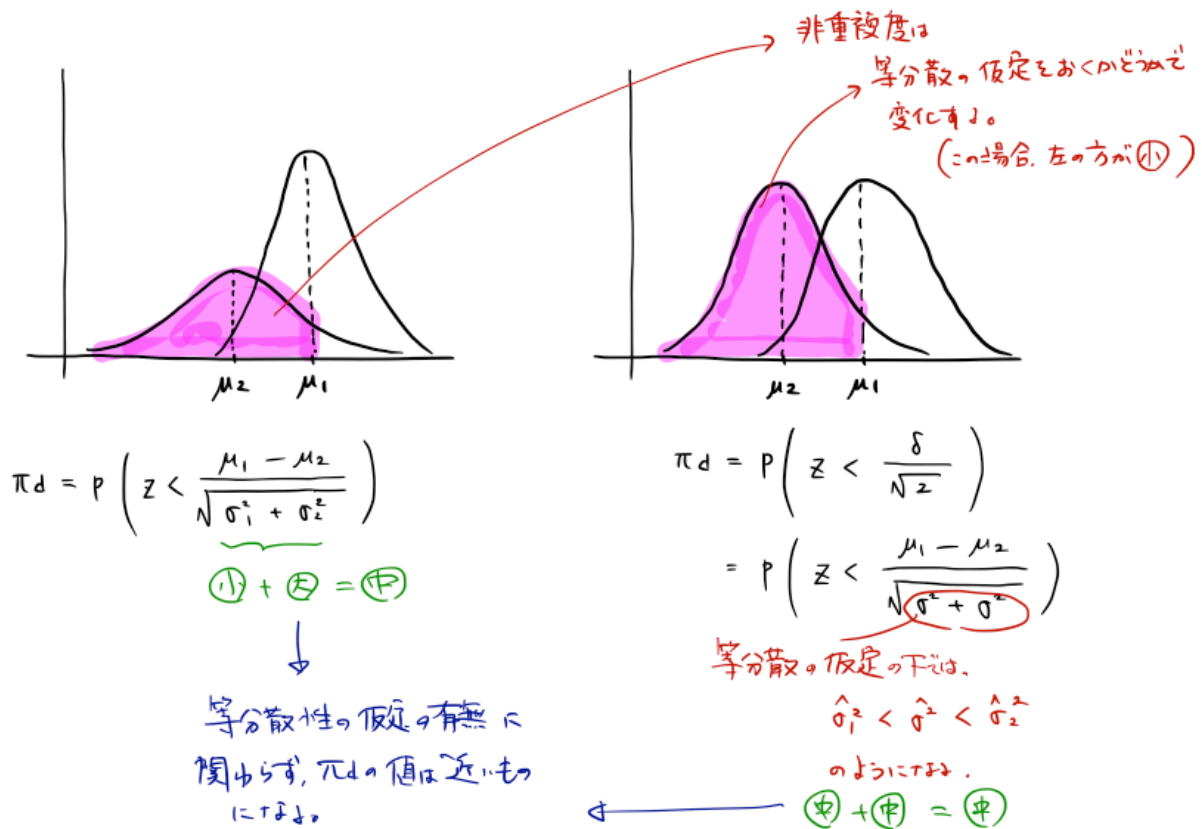


Figure 3.2: 標準偏差の等値制約と非重複度・優越率の変化

メモ

P.81 にあるように、母数の事前分布に一樣分布 $U(0, 100)$ などを stan で設定するにはどのようにしたらよいのだろうか？

メモここまで

stan のコード

```
data{
  int<lower=0> n1;
  int<lower=0> n2;
  vector[1] x1[n1];
  vector[1] x2[n2];
}

parameters{
  real mu1;
  real mu2;
  real<lower=1e-10> sigma1;
  real<lower=1e-10> sigma2;
}

model{
  for(i in 1:n1){
    x1[i]~normal(mu1,sigma1);
  }
  for(i in 1:n2){
    x2[i]~normal(mu2,sigma2);
  }
}

generated quantities{
  real<lower=0,upper=1> p_mu1over2;    //[第 1 群の平均値>第 2 群の平均値] である確率
  real diff_mu;                       //[第 1 群の平均値-第 2 群の平均値]
  real delta;                         //[効果量 (第 2 群の SD を分母とする)]
  real<lower=0,upper=1> p_delta_overc; //効果量>0.5 となる確率
  real<upper=1> u3;                   //[非重複率 (第 1 群の平均が第 2 群の何%点か)]
  real<lower=0,upper=1> p_u3_overc;   //[非重複率>0.8 となる確率]
  real<upper=1> pd;                   //[優越率]
  real<lower=0,upper=1> p_pd_overc;   //[優越率>0.8 となる確率]
  real<upper=1> pbt;                  //[閾上率 (閾値を 5 点とする)]
  real<lower=0,upper=1> p_pbt_overc;  //[閾上率>0.8 となる確率]
  p_mu1over2 = step(mu1-mu2);
  diff_mu = mu1-mu2;
  delta = (mu1-mu2)/sigma2;
  p_delta_overc = step(delta-0.5);
  u3 = normal_cdf(mu1,mu2,sigma2);
  p_u3_overc = step(u3-0.8);
  pd = normal_cdf((mu1-mu2)/sqrt(pow(sigma1,2)+pow(sigma2,2)),0,1);
  p_pd_overc = step(pd-0.8);
  pbt = normal_cdf((mu1-mu2-5)/sqrt(pow(sigma1,2)+pow(sigma2,2)),0,1);
  p_pbt_overc = step(pbt-0.8);
}
```


R のコード

```
x1<-c(
  56,55,55,62,54,63,47,58,56,56,57,52,53,50,50,57,57,55,60,65,53,43,60,51,52,
  60,54,49,56,54,55,57,53,58,54,57,60,57,53,61,60,58,56,52,62,52,66,63,54,50)
x2<-c(
  33,37,59,41,42,61,46,25,32,35,55,44,45,41,33,61,46,16,48,34,27,37,28,31,32,
  20,50,42,26,55,45,36,51,51,50,48,47,39,36,35,32,38,25,66,54,27,35,34,49,39)
n1<-length(x1);n2<-length(x2)
mean(x1);sd(x1);var(x1);summary(x1)
mean(x2);sd(x2);var(x2);summary(x2)
data01 <- list(n1=n1,n2=n2,
               x1=structure(.Data=x1,.Dim=c(n1,1)),x2=structure(.Data=x2,.Dim=c(n2,1)))
scr <- "toyoda2016-exer03.stan"
par <- c("mu1","mu2","sigma1","sigma2","p_mu1over2","diff_mu",
        "delta","p_delta_overc","u3","p_u3_overc","pd","p_pd_overc","pbt","p_pbt_overc")
war <- 1000
ite <- 11000
see <- 1234
dig <- 3
cha <- 4
fit01 <- stan(file=scr,model_name=scr,data=data01,pars=par,verbose=F,seed=see,
              chains=cha,warmup=war,iter=ite)
print(fit01,pars=par,digits_summary=dig)
plot(fit01,pars=par)
traceplot(fit01,inc_warmup=T,pars=par)
```

実行結果

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu1	55.761	0.003	0.671	54.452	55.314	55.760	56.206	57.090	40000	1
mu2	40.378	0.008	1.624	37.176	39.299	40.373	41.452	43.584	40000	1
sigma1	4.699	0.002	0.493	3.853	4.352	4.659	5.002	5.778	40000	1
sigma2	11.443	0.006	1.194	9.387	10.600	11.347	12.184	14.045	40000	1
p_mu1over2	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	40000	NaN
diff_mu	15.383	0.009	1.758	11.939	14.212	15.386	16.544	18.844	40000	1
delta	1.359	0.001	0.208	0.960	1.218	1.356	1.496	1.775	40000	1
p_delta_overc	1.000	0.000	0.005	1.000	1.000	1.000	1.000	1.000	40000	1
u3	0.908	0.000	0.034	0.831	0.888	0.912	0.933	0.962	40000	1
p_u3_overc	0.995	0.000	0.073	1.000	1.000	1.000	1.000	1.000	35283	1
pd	0.891	0.000	0.033	0.816	0.871	0.895	0.915	0.946	40000	1
p_pd_overc	0.989	0.001	0.104	1.000	1.000	1.000	1.000	1.000	32356	1
pbt	0.798	0.000	0.045	0.703	0.770	0.801	0.829	0.877	40000	1
p_pbt_overc	0.508	0.002	0.500	0.000	0.000	1.000	1.000	1.000	40000	1

4 対応のある 2 群のデータ

4.2 3つの視点の区別

(1) 独立した 2 群の群間差 処置あり群と、それとは別の人からなる処置なし群との集団としての差異に関心がある。

(2) 対応のある 2 群の群間差 処置前の群と、同一の人たちから構成される参加後の群との集団としての差異に関心がある。

(3) 対応のある 2 群の個人内差 処置前から処置後にかけての個人内変化（差得点）の平均と個人差に関心がある¹¹。

4.3 相関を考慮した個人内変化の分析

処置前得点と処置後得点の相関

- ほぼゼロ、マイナス：不安で受けられない。
- 高すぎる：伸びしろがなくつまらない。頑張った分の伸びしろが欲しい。

メモ：処置前から処置後への変化と相関係数について

なぜ、処置前値と処置後値の相関がゼロや負では、「不安で受けられない」のか？単純に、「相関がゼロ = 受けても効果なし」、「相関が負 = 受けたら逆効果」、と評価してしまいそうだが、本当にそうか？

まず、処置前処置後得点の標準偏差が σ と共通であるとする。モデル分布が正規分布であるとする、差得点の分布は以下のようになる。

$$x_{1i}^* - x_{2i}^* \sim N(\mu_1 - \mu_2, \sigma\sqrt{2(1-\rho)}) \quad (4.22)$$

これによると、平均的に効果があるか（あるいは逆効果か）どうかは $\mu_1 - \mu_2 > 0$ かどうかの問題であって、相関 ρ の問題ではないことが分かる。相関 ρ は、差得点の SD しか左右していない。むしろ、 ρ が高いほど差得点の SD は小さくなるわけで、「その処置の効果には個人差が少ない」ことになる。

一方、

$$Con = p((x_{1i} - x_{1j})(x_{2i} - x_{2j}) > 0) = 0.5 + \frac{1}{\pi} \sin^{-1}(\rho) \quad (4.35)$$

で定義される同順率は、 $\rho = 0$ のとき $Con = 0.5$ となり、 $\rho < 0$ では $Con < 0.5$ となる。つまり、 $\rho = 0$ は処置前から処置後にかけて集団内での相対的な順位が入れ替わる確率が 5 割を切る点として解釈できる。処置前値と処置後値の相関がゼロや負では「不安で受けられない」というのは、絶対的な得点の変化に関する議論ではなく、相対的な順位の入れ替わりを巡る議論であるということになる。

同様の問題は、交差遅延効果モデルや同時効果モデルにより「経時変化」について議論する際にも当てはまる。

メモここまで

差得点の SD 処置前と処置後得点の標準偏差が共通 σ であるとき、差得点の SD $\sigma' = \sigma\sqrt{2(1-\rho)}$ とする。

4.3.3 差得点の効果量

$$\delta' = \frac{\mu_1 - \mu_2}{\sigma'} = \frac{\mu_1 - \mu_2}{\sigma\sqrt{2(1-\rho)}} \quad (4.25)$$

4.3.4 差得点の優越率

$$\pi'_d = p(x_{1i}^* - x_{2i}^* > 0) = p\left(z < \frac{\mu_1 - \mu_2}{\sigma'}\right) = p\left(z < \frac{\mu_1 - \mu_2}{\sigma\sqrt{2(1-\rho)}}\right) \quad (4.28)$$

¹¹ 本文にはこのように明確に言及されているわけではない。

4.3.5 差得点の閾上率

$$\pi'_c = p(x_{1i}^* - x_{2i}^* > c) = p\left(z < \frac{\mu_1 - \mu_2 - c}{\sigma'}\right) = p\left(z < \frac{\mu_1 - \mu_2 - c}{\sigma\sqrt{2(1-\rho)}}\right) \quad (4.31)$$

4.3.7 同順率

$$Con = p((x_{1i} - x_{1j})(x_{2i} - x_{2j}) > 0) = 0.5 + \frac{1}{\pi} \sin^{-1}(\rho) \quad (4.35)$$

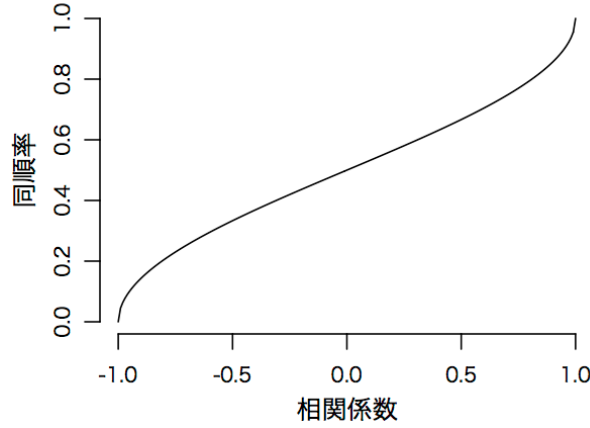


Figure 4.1: 相関係数と同順率

4.4 標準偏差が異なる 2 変量正規分布モデル

処置前と処置後の SD を σ_1, σ_2 とすると、差得点の SD $\sigma' = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$ となる。

5 実験計画による多群の差の推測

5.1 独立した 1 要因の推測

5.1.1 独立した 1 要因モデル

$$y_{ij} = \mu_j + e, \quad e \sim N(0, \sigma_e) \quad (5.1)$$

5.1.2 全平均と水準の効果

全平均 人種など構成比率が分かっているなど特別な場合を除いては、単純平均として生成。

$$\mu = \frac{1}{a}(\mu_1 + \cdots + \mu_a) \quad (5.8)$$

母数は n_j を含んで定義されるものではないから、 n_j に比例した重みづけ平均を取る必要はない。

水準の効果

$$a_j = \mu_j - \mu \quad (5.9)$$

全水準の合計はゼロとなる。

5.1.5 要因の効果の大きさ

水準の効果 a_j が誤差変数 e と独立であるとする、

$$\sigma_y^2 = \sigma_a^2 + \sigma_e^2 \quad (5.13)$$

と分解できる。

説明率

$$\eta^2 = \frac{\sigma_a^2}{\sigma_y^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (5.15)$$

効果量

$$\delta = \frac{\sigma_a}{\sigma_e} \quad (5.16)$$

5.1.7 連言命題が正しい確率

研究仮説「 $\mu_4 > \mu_3 > \mu_1 > \mu_2$ 」が正しい確率は、生成量

$$u_{\mu_4 > \mu_3}^{(t)} \times u_{\mu_3 > \mu_1}^{(t)} \times u_{\mu_1 > \mu_2}^{(t)} \quad (5.19)$$

の EAP で評価できる。

研究仮説「 $\mu_4 > (\mu_3, \mu_1) > \mu_2$ 」が正しい確率は、生成量

$$u_{\mu_4 > \mu_3}^{(t)} \times u_{\mu_4 > \mu_1}^{(t)} \times u_{\mu_3 > \mu_2}^{(t)} \times u_{\mu_1 > \mu_2}^{(t)} \quad (5.20)$$

の EAP で評価できる¹²。

5.1.8 特に興味のある 2 水準間の推測

2 群のデータだけ用いて再分析しないこと。ベイズ統計学は、データの 2 度使いをしない。

全データを使って求めた $\mu_j^{(t)}$ と $\sigma_e^{(t)}$ から、効果量、非重複度、優越率、閾上値を算出する。

5.2 独立した 2 要因の推測

5.2.1 独立した 2 要因モデル

$$y_{ijk} = \mu + a_j + b_k + (ab)_{jk} + e, \quad e \sim N(0, \sigma_e) \quad (5.26)$$

ただし、

$$\sum_{j=1}^a a_j = 0, \quad \sum_{k=1}^b b_k = 0, \quad \sum_{j=1, k}^a (ab)_{jk} = 0, \quad \sum_{j, k=1}^b (ab)_{jk} = 0 \quad (5.27)$$

とする。結局、 μ が 1 つ、 a_j が $a - 1$ 個、 b_k が $b - 1$ 個、 $(ab)_{jk}$ が $(a - 1)(b - 1)$ 個、 σ_e が 1 個（合計 $ab + 1$ 個）の母数が推測される。

例えば 3×2 の 2 要因の場合、以下の母数（セル平均ではないので注意）のうち太字の部分以外は推測されない。

¹² $p(\mu_4 > \mu_3 > \mu_2 \mid \mathbf{y})$ と $p(\mu_4 > \mu_1 > \mu_2 \mid \mathbf{y})$ の和としてしまいそうなので注意。

	b_1	b_2
a_1	$(ab)_{11}$	$(ab)_{12}$
a_2	$(ab)_{21}$	$(ab)_{22}$
a_3	$(ab)_{31}$	$(ab)_{32}$

それ以外の母数については、

$$\begin{aligned}
a_3 &= -(a_1 + a_2) \\
b_2 &= -b_1 \\
(ab)_{31} &= -((ab)_{11} + (ab)_{21}) \\
(ab)_{12} &= -(ab)_{11} \\
(ab)_{22} &= -(ab)_{21} \\
(ab)_{32} &= -(ab)_{31}
\end{aligned}$$

に対応する生成量を算出することで評価する。

5.2.3 要因の効果の大きさ

1 要因のときと同様、

$$\sigma_y^2 = \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \sigma_e^2 \quad (5.39)$$

と分解できるとする。ただし、

$$\begin{aligned}
\sigma_a^2 &= \frac{1}{a} \sum_{j=1}^a a_j^2 \\
\sigma_b^2 &= \frac{1}{b} \sum_{k=1}^b b_k^2
\end{aligned} \quad (5.40)$$

$$\sigma_{ab}^2 = \frac{1}{ab} \sum_{j=1}^a \sum_{k=1}^b (ab)_{jk}^2 \quad (5.41)$$

である。

説明率

$$\eta_a^2 = \frac{\sigma_a^2}{\sigma_y^2}, \quad \eta_b^2 = \frac{\sigma_b^2}{\sigma_y^2}, \quad \eta_{ab}^2 = \frac{\sigma_{ab}^2}{\sigma_y^2}, \quad \eta_t^2 = \frac{\sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2}{\sigma_y^2} \quad (5.42)$$

効果量

$$\delta_a = \frac{\sigma_a}{\sigma_e}, \quad \delta_b = \frac{\sigma_b}{\sigma_e}, \quad \delta_{ab} = \frac{\sigma_{ab}}{\sigma_e} \quad (5.43)$$

5.2.4 セル平均の事後分布

生成量

$$\mu_{jk}^{(t)} = \mu^{(t)} + a_j^{(t)} + b_k^{(t)} + (ab)_{jk}^{(t)} \quad (5.44)$$

によって、近似する。

5.2.5 特に興味のある 2 セル間の比較

セル平均と誤差 SD の生成量 ($\mu_{jk}^{(t)}$ と $\sigma_e^{(t)}$) を利用して、平均の差、効果量、非重複度、優越率、閾上率を算出する。

5.3 章末問題

(1) 独立した 1 要因

R コード (stan コードも R で書く)

```
x1 <- c(05.02, 06.67, 08.17, 02.79, 08.13, 06.34, 06.32, 03.97); n1 <- length(x1)
x2 <- c(09.89, 09.58, 11.20, 09.05, 12.33, 09.39, 10.88, 09.37, 17.40); n2 <- length(x2)
x3 <- c(10.20, 07.29, 07.57, 03.42, 05.82, 10.92, 05.21, 13.47, 08.64, 06.05); n3 <- length(x3)
data01 <- list(n1=n1,n2=n2,n3=n3,x1=x1,x2=x2,x3=x3)
```

```
stancode <- '
data{
  int<lower=0> n1;
  int<lower=0> n2;
  int<lower=0> n3;
  real x1[n1];
  real x2[n2];
  real x3[n3];
}
parameters{
  real mu1;
  real mu2;
  real mu3;
  real<lower=0> sigma;
}
model{
  x1~normal(mu1,sigma);
  x2~normal(mu2,sigma);
  x3~normal(mu3,sigma);
}
'

fit01 <- stan(model_code=stancode,iter=10000,warmup=1000,chains=4,data=data01)
print(fit01,digits_summary=3)
```

実行結果

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu1	5.925	0.005	0.988	3.971	5.286	5.935	6.561	7.901	36000	1
mu2	11.009	0.005	0.923	9.205	10.411	11.004	11.608	12.850	36000	1
mu3	7.855	0.005	0.871	6.124	7.281	7.858	8.429	9.572	36000	1
sigma	2.744	0.002	0.426	2.067	2.442	2.694	2.987	3.728	30738	1
lp__	-38.935	0.012	1.533	-42.858	-39.685	-38.593	-37.809	-37.025	15096	1

(2) 独立した 2 要因

rstan コード (動かず)

```
A <- c(rep(1,49),rep(2,49))
B <- c(rep(1,10),rep(2,8),rep(3,7),rep(4,9),rep(5,8),rep(6,7),
      rep(1,10),rep(2,8),rep(3,7),rep(4,9),rep(5,8),rep(6,7))
y <- c(140,146,149,136,147,147,143,143,143,141,
      139,136,136,140,135,132,140,134,
      123,127,131,130,138,128,129,
      115,120,118,118,121,124,129,119,128,
      128,124,123,121,122,126,131,122,
      121,121,120,116,117,113,118,
```

```

143,141,142,145,149,145,143,141,142,155,
138,134,142,136,135,136,131,133,
131,128,128,128,127,130,130,
117,125,132,122,119,122,129,117,127,
117,120,124,122,122,122,118,122,
119,125,122,116,119,113,122)
a <- length(unique(A))
b <- length(unique(B))
N <- length(y)
data01 <- list(y=y,A=A,B=B,a=a,b=b,N=N)

stancode <- '
data{
  int<lower=1> N; int<lower=1> A[N]; int<lower=1> B[N];
  vector<lower=0>[N] y; int<lower=2> a; int<lower=2> b;
}
parameters{
  real mu;
  vector[a-1] ef_a; vector[b-1] ef_b; matrix[a-1,b-1] ef_ab;
  real<lower=0> sigma;
}
transformed parameters{
  vector[a] ef_a_all; vector[b] ef_b_all; matrix[a,b] ef_ab_all; // 主効果と交互作用全体
  for(k in 1:(b-1)){ ef_a_all[k] = ef_a[k]; } ef_a_all[b] = -sum(ef_a);
  for(j in 1:(a-1)){ ef_b_all[j] = ef_b[j]; } ef_b_all[a] = -sum(ef_b);
  for(j in 1:(a-1)){ for(k in 1:(b-1)){ ef_ab_all[j,k] = ef_ab[j,k]; } }
  for(k in 1:(b-1)){ ef_ab_all[a,k] = -sum(col(ef_ab,k)); }
  for(j in 1:(a-1)){ ef_ab_all[j,b] = -sum(row(ef_ab,j)); }
  ef_ab_all[a,b] = -sum(sub_row(ef_ab_all,a,1,b-1));
}
model{
  for(i in 1:N){
    y[i] ~ normal(mu+ef_a_all[A[i]]+ef_b_all[B[i]]+ef_ab_all[A[i],B[i]],sigma);
  }
}
,

fit01 <- stan(model_code=stancode,iter=10000,warmup=1000,chains=4,data=data01)
print(fit01,digits_summary=3)

```

6 比率とクロス表の推測

6.3 比率の推測 II (1つの多項分布)

i 番目のカテゴリの仮の確立の事前分布を

$$\ddot{p}_i \sim N(0, 1) \quad (6.25)$$

とし、

$$p_i = \frac{\ddot{p}_i}{\sum_{i=1}^k \ddot{p}_i} \quad (6.26)$$

と母数を構成する¹³。

¹³ $\sum_{i=1}^k p_i = 1$ でなくてはならない。

6.5 対応のあるクロス表の推測（1つの多項分布に構造が入る）

6.5.1 2×2 のクロス表の推測

Table 6.12: 対応のある 2×2 のクロス表

	B_1	B_2	計
A_1	x_{11}	x_{12}	$x_{1.}$
A_2	x_{21}	x_{22}	$x_{2.}$
計	$x_{.1}$	$x_{.2}$	n

Table 6.13: 出現確率の比率

	B_1	B_2	計
A_1	p_{11}	p_{12}	$p_{1.}$
A_2	p_{21}	p_{22}	$p_{2.}$
計	$p_{.1}$	$p_{.2}$	1.0

このとき、データ \mathbf{x} 、母数 \mathbf{p} の尤度は、多項分布

$$f(\mathbf{x} | \mathbf{p}) = f(x_{11}, x_{12}, x_{21}, x_{22} | p_{11}, p_{12}, p_{21}, p_{22}) \quad (6.53)$$

で表現できる。

事前分布については、 \check{p}_{ij} が仮に

$$\check{p}_{ij} \sim U(0, 1) \quad (6.54)$$

であるとし、和が1となるよう

$$p_{ij} = \frac{\check{p}_{ij}}{\check{p}_{11} + \check{p}_{12} + \check{p}_{21} + \check{p}_{22}} \quad (6.55)$$

と構成する。

6.5.2 独立と連関

変数 A のカテゴリ A_i と変数 B のカテゴリ B_j について、

$$p_{ij} \neq p_{i.} \times p_{.j} \quad (6.62)$$

であれば、連関している（独立でない）ことになる。

- すべての i, j について $p_{ij} = p_{i.} \times p_{.j}$ であれば、2つの変数は独立である。
- 少なくとも1つの組について $p_{ij} \neq p_{i.} \times p_{.j}$ であれば、2つの変数は独立でない（連関している）。

6.5.3 ピアソン残差・クラメルの連関係数（生成量）

ピアソン残差 独立であれば0となり、独立な場合より高い比率で観測されていれば正、独立な場合より低い比率で観測されていれば負の値をとる指標

$$e_{ij} = \frac{p_{ij} - p_{i.}p_{.j}}{\sqrt{p_{i.}p_{.j}}} \quad (6.65)$$

クラメルの連関係数 クロス表全体での連関の程度を示す指標

$$V = \sqrt{e_{11}^2 + e_{12}^2 + e_{21}^2 + e_{22}^2} \quad (6.66)$$

6.5.4 $a \times b$ のクロス集計表の推測

事前分布については、 \check{p}_{ij} が仮に

$$\check{p}_{ij} \sim U(0, 1) \quad (6.77)$$

であるとし、和が1となるよう

$$p_{ij} = \frac{\check{p}_{ij}}{\check{p}_{11} + \cdots + \check{p}_{1j} + \cdots + \check{p}_{ij} + \cdots + \check{p}_{aj} + \cdots + \check{p}_{ab}} \quad (6.78)$$

と構成する。

クラメルの連関係数は、

$$V = \sqrt{\frac{e_{11}^2 + \cdots + e_{1j}^2 + \cdots + e_{1b}^2 + \cdots + e_{a1}^2 + \cdots + e_{aj}^2 + \cdots + e_{ab}^2}{\min(a, b) - 1}} \quad (6.80)$$

6.5.5 連言命題が正しい確率

ピアソン残差 e_{ij} を用いて、例えば生成量

$$u_{e_{11}>0}^{(t)} \times u_{e_{33}>0}^{(t)} \times u_{e_{31}<0}^{(t)} \times u_{e_{13}<0}^{(t)} \quad (6.81)$$

の EAP を算出する。

Q & A

Q3 データ数 n が大きいとき、有意性検定とベイズ推測ではどのような違いがありますか。

有意性検定では、 n が大きくなると p は平均的にいくらでも 0 に近づく。ビッグデータに対しては、有意性検定は無効。検定力分析によるサンプルサイズ n の制限・設計は纏足と同じ。有意性検定の制度を守るために、それに合わせて n を制限・設計することは本末転倒。

Q5 $n-1$ で割ったりしないのですか？

不偏性とは「統計量の期待値を取ると、それが（未知ではあっても固定された）母数に一致する」という性質。ベイズ統計学では、母数は固定された値ではなく確率分布するので、通常の意味での不偏性という概念は存在しなくなる。

Q6 実験データの分析はアンバランスの時に計算方法が異なるのでは？

ベイズ的アプローチでは、アンバランスか否かで計算手順は変わらない。伝統的統計学における分散分析では、その実験でたまたまとったデータの数、調べたい効果を決める¹⁴。こう定義しないと F 値が導出できないからだが、分析の手段が分析の目的を規定しているという本末転倒なことが起こっている。

Q12 事後分布や推定値は式では表せないのですか？

事後分布は、よく知られた理論分布に一致することもある。ただし、そのような公式化は不可能であつたり煩雑であつたりする。また、公式化が可能であるように事前分布を選ぶ必要が出てくる。これは、母数に対する純粋な信念に基づいて事前分布を選ぶという本来の趣旨から誤っている。

¹⁴要因の効果を算出する時にサンプルサイズで重みづけするから。