練習問題・演習問題解答例

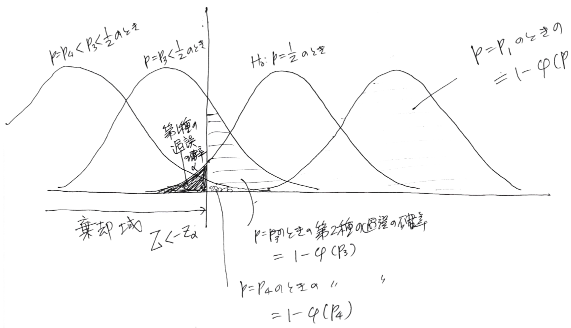
7.1 男の子の人数 X の期待値と分散を $\mathrm{E}[X]=\mu(=np)$, $\mathrm{Var}(X)=\sigma^2(=np(1-p))$ とする．3.2 節の期待値，分散の性質を用いると

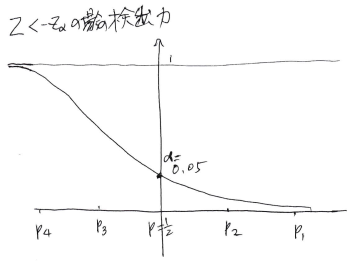
$$\mathrm{E}[Z]=\mathrm{E}[\frac{X-\mu}{\sigma}]=\frac{1}{\sigma}\mathrm{E}[X-\mu]=\frac{1}{\sigma}\left(\mathrm{E}[X]-\mu\right)=0$$

$$\begin{aligned}\mathrm{Var}[Z]&=\mathrm{E}[(Z-\mathrm{E}[Z])^2]=\mathrm{E}[Z^2]=\mathrm{E}[\frac{(X-\mu)^2}{\sigma^2}]=\frac{1}{\sigma^2}\mathrm{E}[(X-\mu)^2]\\&=\frac{1}{\sigma^2}\mathrm{Var}[X]=\frac{1}{\sigma^2}\sigma^2=1\end{aligned}$$

であり，平均が 0，分散は 1 になることがわかる．

7.2 棄却域を下図のようにとると $p=p_3,p_4$ となるにつれて第 2 種の過誤の確率は小さくなり，検出力が大きくなる．逆に $p=p_1$ のように，真の分布が $H_0:p=\frac{1}{2}$ に比べて棄却域から遠くなると，第 2 種の過誤の確率は大きくなり，検出力が小さくなる．





7.4 (1) 帰無仮説 H_0: $p = 0.5$, 対立仮説 H_1: $p > 0.5$

(2) $P(X = k) = {}_{144}C_k \left(\frac{1}{2}\right)^1 44$ であり，

$$E[X] = np = 144 \times \frac{1}{2} = 72, \quad Var[X] = np(1-p) = 144 \times \left(\frac{1}{2}\right)^2 = 36$$

(3) 中心極限定理により，帰無仮説 H_0 の下で，X は正規分布 $N(72, 36)$ に近づく．

(4) X を標準化した分布なので，標準正規分布 $N(0, 1)$ に近づく．

(5) 有意水準 5%の棄却域は，$Z > 1.64$ であり，標本から計算された Z の値は，$Z = \frac{80-72}{\sqrt{36}} = \frac{4}{3} = 1.333$ であるから，帰無仮説は棄却されない．したがって，有意水準 5%でチーム A の勝つ確率は 5 割を超えているとまでは言えない．

8.1 練習問題の直前の説明どおり `t.test` を用いて計算すればよい．(省略)

8.2 μ_1, μ_2 をそれぞれ熊谷と東京の最高気温の期待値とすると，帰無仮説と対立仮説はそれぞれ，$H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 > \mu_2$ であり，片側検定である．熊谷の最高気温と東京の最高気温の差は
1.3, 1.5, 4.3, 2.3, 2.0, 4.0, 2.4, 2.1, 0.8, 2.2, -0.1, 2.0, 3.0, 2.4, 1.6, 2.4, 2.8, 2.3, 0.9, 1.4, 0.7, 2.5, 0.2, 0.9, 2.1, -0.8, 1.3, 1.3, -2.0, 2.3, 2.4
であり，このデータを x とすると，

```
> mean(x)
[1] 1.693548
> tvalue=mean(x)/(sqrt(var(x)/31))
> tvalue
[1] 7.414544
> pt(tvalue, df=31, lower.tail=F)
[1] 1.188462e-08
```

となり，p 値が 1.188×10^{-8} と非常に小さい値となり，帰無仮説 $H_0 : \mu_1 = \mu_2$(熊谷の最高気温と東京の最高気温は変わらない) は棄却される．

熊谷と東京の気温をそれぞれ x，y として，t.test(x, y, paired=T) でも検定できる．

8.3 熊谷と多治見の気温を R に下記のように入力し，

```
x<-c(34.8, 36.4, 37.1, 32.7, 35.9, ... , 34.2, 33.8) #熊谷

y<-c(36.3, 33.7, 28.7, 26.0, 35.1, ... , 34.7, 33.3) #多治見
```

対でない 2 標本 x，y の両側検定を Student の t 検定を用いて行うには，
`t.test(x, y, alternative="two.sided", paired=F, var.equal=T)`
とすればよい．この結果，$T = -0.4805$ となり，その p 値は，$\mathrm{P}\,(|T| > -0.4805) = 0.6326 > 0.05$ であり，有意水準 5%で帰無仮説は棄却できない．すなわち，熊谷と多治見の最高気温に差があるとは言えない．

一方，対標本として，
`t.test(x, y, alternative="two.sided", paired=T)`
とした場合には，8.2 節でみたように $\mathrm{P}\,(|T| > -0.6784) = 0.5027$ (自由度 30) となり，対標本の場合のほうが p 値が少し小さくなり，より検出力の高い検定ができる．

8.4 (1) 訂正された問題の解答例を記載します．

ボルトの直径の誤差の母平均を μ とすると，偏りがないときには $\mu = 0$ であるから，帰無仮説 $H_0 : \mu = 0$, 対立仮説 $H_1 : \mu \neq= 0$ である．

8.1 節の t-検定の手法に従って，R で検定を行うと，

```
> x=c(4.8, -3.6, 5.2, -0.4, 3.9, -4.9, -5.3, 1.6, -2.4, 4.1)
> t.test(x, mu=0, alternative=''two.sided'')

        One Sample t-test
data:  x
t = 0.22879, df = 9, p-value = 0.8241
```

```
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -2.666216  3.266216
sample estimates:
mean of x
      0.3
```

となり，p-値は充分大きく母平均が 0(ミクロン) から偏っているとは言えない．

(2) 省略．8.2 節と同様に検定を行え．

9.1 R で，

```
> pairwise.t.test(hightemp$temp, hightemp$city, p.adj="holm")
```

を実行すると，

	E	Ko	Ku	T
Ko	1.00000	-	-	-
Ku	1.00000	1.00000	-	-
T	0.99088	0.99088	0.67692	-
Y	0.02035	0.01704	0.04457	0.00024

となる．このことより，山形 (Y) と他の 4 都市の間の p 値がいずれも 0.05 より小さく，山形と他の都市との間には有意な差があり，他の 4 都市には差があるとは言えないことがわかる．

9.5 `warpbreaks` は，毛糸の紡績の際に，毛糸の種類 A, B と糸の張りの強さ (弱 (L), 中 (M), 強 (H)) によって，糸が切れる回数がどのように違うかを調べたデータであり，R で，

```
> warpbreaks
   breaks wool tension
1      26    A       L
2      30    A       L
```

```
...        ...    ...
54     28    B       H
>
```

とすると，データが見える．このデータに 2 元配置のモデルを適用すると，下記のように張りの強さが最も強く影響しており，毛糸と張りの強さの交互作用も影響があるという結果が得られた．

```
> summary(fm1<-aov(breaks~wool+tension+wool+wool*tension, data=warpbreaks))
             Df Sum Sq Mean Sq F value   Pr(>F)
wool          1    451   450.7   3.765 0.058213 .
tension       2   2034  1017.1   8.498 0.000693 ***
wool:tension  2   1003   501.4   4.189 0.021044 *
Residuals    48   5745   119.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

さらに，張りの強さについて Tukey の HSD による多重比較を行うと，下記のように強 (H) と中 (M) には余り差がなく，弱 (L) の場合には，他の 2 つの場合に比べて，切れる回数が多いことがわかる．

```
> TukeyHSD(fm1, "tension")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = breaks ~ wool + tension + wool + wool * tension,
          data = warpbreaks)

$tension
          diff       lwr       upr     p adj
M-L -10.000000 -18.81965 -1.180353 0.0228554
H-L -14.722222 -23.54187 -5.902575 0.0005595
```

```
H-M  -4.722222 -13.54187  4.097425 0.4049442

> plot(TukeyHSD(fm1, "tension")) #図の表示は省略
```

10.1 標本 $x_1, x_2, \ldots, x_n$ が得られる密度は

$$f(x_1) \times f(x_2) \times \cdots \times f(x_n) = \lambda^n e^{-\lambda(x_1+\cdots+x_n)}$$

であるから，対数尤度関数は，

$$\ell(\lambda) = n \log \lambda - \lambda(x_1 + \cdots + x_n)$$

よって，最尤推定値は

$$\frac{d\ell(\lambda)}{d\lambda} = n\frac{1}{\lambda} - (x_1 + \cdots + x_n) = 0$$

の解であり，$\widehat{\lambda} = \frac{n}{x_1+\cdots+x_n}$ である．

10.2 図 10.2 の接線の式 $y - g(x_0) = g'(x_0)(x - x_0)$ と x 軸との交点は $y = 0$ とすると求まり，$-g(x_0) = g'(x_0)(x - x_0)$ の解であるから，$x_1 = x_0 - \frac{g(x_0}{g'(x_0)}$．これを繰り返すので，$x_2 = x_1 - \frac{g(x_1}{g'(x_1)}, \ldots, x_n = x_{n-1} - \frac{g(x_{n-1})}{g'(x_{n-1})}$ となる．

10.5 尤度関数は

$$\begin{aligned} L(\lambda) &= \frac{\lambda^{k_1}}{k_1!}e^{-\lambda} \times \frac{\lambda^{k_2}}{k_2!}e^{-\lambda} \times \frac{\lambda^{k_n}}{k_n!}e^{-\lambda} \\ &= \frac{\lambda^{k_1+\cdots+k_n}}{k_1!\cdots k_n!}e^{-n\lambda} \end{aligned}$$

であり，対数尤度関数は

$$\ell(\lambda) = \log L(\lambda) = (k_1 + \cdots + k_n) \log \lambda - \log(k_1! \cdots k_n!) - n\lambda$$

よって，

$$0 = \frac{d\ell}{d\lambda} = \frac{k_1 + \cdots + k_n}{\lambda} - n$$

ゆえに，$\hat{\lambda} = \frac{k_1+\cdots+k_n}{n}$ である．

14.1 $g(x)$ として $[0, 1]$ 区間の一様分布を用いる．このとき，$f'(x) = x^2(1 - x)(3 - 5x)$ であり，$f(x)$ は $x = \frac{3}{5}$ で最大値 $c = 60\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^2 = \frac{1296}{625}$ をとる．よって，乱数発生手順は，

(i) $[0, 1]$ 区間の一様乱数 X と U を生成する．

(ii) $U \leq \frac{f(X)}{cg(X)}$ を満たすならば，X を i 番目の乱数 X_i とする．そうでなければ，X と U を廃棄して，1. に戻る．

(iii) 上記のプロセスを n 個の乱数 $X_1, \ldots, X_n$ が得られるまで繰り返す．

14.2 $\mathrm{E}[Y] = \mathrm{E}[aX+b] = a\mathrm{E}[X]+b = b$, $\mathrm{Var}[Y] = \mathrm{Var}[aX+b] = a^2\mathrm{Var}[Y] = a^2$ であるから，Y は $\mathrm{N}(b, a^2)$ にしたがう．よって，$b = \mu$, $a = \sigma$ とすれば，$Y = \sigma X + \mu$ により，標準正規分布に従う X から $\mathrm{N}(\mu, \sigma^2)$ に従う Y が得られる．

14.3 R のプログラムは，下記のとおりである．実際に R で実行してみよ．

```
u1=runif(1000)
u2=runif(1000)
x1=sin(2*pi*u1)*sqrt(-2*log(u2))
x1=sin(2*pi*u1)*sqrt(-2*log(u2))
x=c(x1, x2)
x=x[abs(x)<4] # ヒストグラムを描くために 2000 個の乱数のうち絶
対値が 4 以上のものを除く
hist(x, breaks=(-20:20)*0.2, freq=F, ylim=c(0,0.45), main="Box-Muller
法")
   # freq=F は相対度数でグラフを描く
par(new=T)
plot(function(x) dnorm(x), -4, 4, main="", ylab="", ylim=c(0, 0.45))
```

14.4 (i) に対する R のプログラムは下記のとおりである．

```
n=100
m=1000
```

```
s=rep(NA, m)
for(i in 1:m){
  x=runif(n)
   s[i]=mean(exp(-x^2))
 }
mean(s)
sd(s)
```

(ii) は，`n=50` とし，`for` 文の中を

```
   x1=runif(n); x2=1-x1; x=c(x1, x2)
   s[i]=mean(exp(-x^2))
```

とすればよい.

例題 14.2　例題 14.2 で用いた一様にばらまかれた乱数の生成法に関するプログラムを掲載しておく.

```
n=100; nsim=1000;p=rep(NA, nsim)
for(i in 1:nsim){
x=runif(n);y=runif(n);
r=x^2+y^2; ncount=length(r[r<1]);
p[i]=4*ncount/n}
mean(p)
sd(p)

m=316; n=m^2; nsim=1000;p=rep(NA, nsim)
for(i in 1:nsim){
x=(runif(m^2)+outer(0:(m-1), rep(1, m), rep))/m
y=(runif(m^2)+t(outer(0:(m-1), rep(1, m), rep)))/m
r=x^2+y^2; ncount=length(r[r<1]);
p[i]=4*ncount/n}
mean(p)
sd(p)
```

```
m=400; n=100
z=rnorm(m^2, mean=0, sd=1/sqrt(2)) +rnorm(m^2, mean=0, sd=1/sqrt(2))*1i
z=matrix(z, c(m, m))
ev=eigen(z)$values
x=Re(ev)/sqrt(m); y=Im(ev)/sqrt(m);
x1=x[(abs(x) <0.5)&(abs(y)<0.5)]
y1=y[(abs(x) <0.5)&(abs(y)<0.5)]
n1=length(x1)
rand=sample(1:n1, n, replace=F)
x=x1[rand]+0.5;y=y1[rand]+0.5
plot(x, y)
```

14.5 求めたい確率は

$$\begin{aligned}
&\int_{60}^{100} c\left(\frac{x}{100}\right)^a\left(1-\frac{x}{100}\right)^b dx \\
&\quad = 100c\int_{0.6}^{1} y^a\,(1-y)^b\,dy \\
&\quad = 40c\int_{0}^{1}(0.4u+0.6)^a(0.4(1-u))^b du
\end{aligned}$$

である．この積分の 2 行目は，不完全ベータ関数と呼ばれる積分であり，積分値を理論的に求めることは困難である．ここでは，14.4.1 節の負の相関法を用いて近似値を求めてみよう．$(0,1)$ 区間の乱数 u を発生させまた，負の相関法を用いるために，$u' = 1-u$ も用いる．

下のプログラムで最初の結果は 2000 個の一様乱数を発生させて，通常のモンテカルロ法を実行した場合であり，後半は 1000 個の一様乱数を用いて負の相関法を用いた場合である．この確率の真の値は，0.655 であり，負の相関法の方が真の値に近い．このプログラムを 100 回程度繰り返して，両者のばらつきを比較すると良い．

```
> a=3.50; b=1.33;
> n=2000
```

```
> u=runif(n)
> Sbar=sum(((0.4*u+0.6)^a)*((0.4*(1-u))^b))
> Sbar=Sbar/n*40*0.3813
> Sbar
[1] 0.6649133
>
> a=3.50; b=1.33;
> ndash=1000;
> u=runif(ndash)
> Sbar=sum(((0.4*u+0.6)^a)*((0.4*(1-u))^b)+((0.4*(1-u)+0.6)^a)*((0.4*u)^b))
> Sbar=Sbar/(2*ndash)*40*0.3813
> Sbar
[1] 0.655981
```