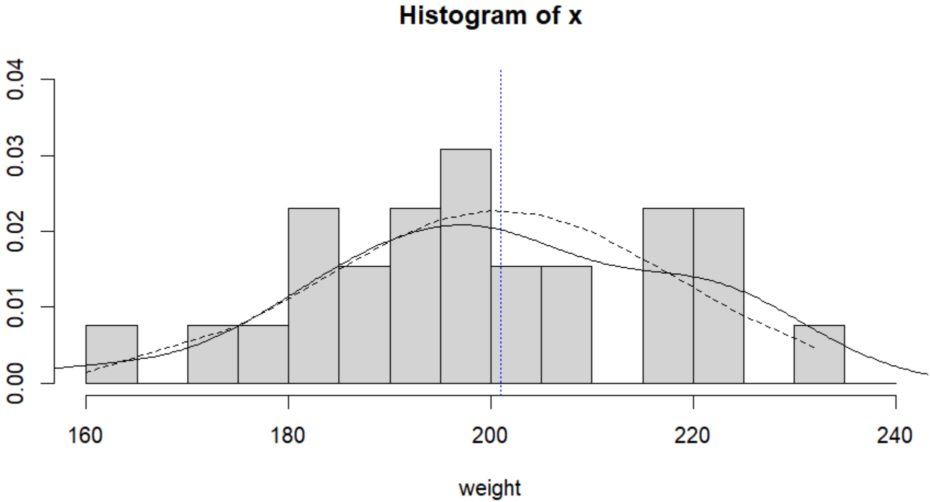
数理统计 week 4

学业辅导中心


Histogram of x
0.00
0.01
0.02
0.03
0.04
160
180
200
220
240
weight

习题 4.1.2

```r
x <- c(160,175,180,185,185,185,190,190,195,195,195,200,200,
       200,200,205,205,210,210,218,219,220,222,225,225,232)
breaks_manual <- seq(160, 240, by = 5)
hist(x, xlab = "weight", ylab = " ", pr = T, ylim = c(0, 0.04), breaks =
    breaks_manual)
lines(density(x))
y <- dnorm(x, mean = mean(x), sd = sd(x))
lines(y~x, lty = 2)
```

根据画出的图可以看出在体重较小的时候正态分布拟合较好, 而大体重正态分布拟合的结果不好. 由于样本量比较小, 也可以认为用正态分布拟合是合理的.

习题 4.1.2

根据 4.1.7 式和 4.1.8 式,

$$
\begin{aligned}
\hat{\mu} &= \bar{X} \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i \\
&= \frac{1}{26}(160 + 175 + \ldots + 225 + 232) \\
&= 201
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \\
&= \frac{1}{26}\left[(160 - 201)^2 + \ldots + (232 - 201)^2\right] \\
&= 293.9231
\end{aligned}
$$

习题 4.1.2

根据定理 6.1.2, 极大似然估计量具有不变性, 因此

$$\begin{aligned}\hat{\sigma} &= \sqrt{\hat{\sigma}^2} \\ &= \sqrt{293.9231} \\ &= 17.14418\end{aligned}$$

$$\begin{aligned}\frac{\hat{\mu}}{\hat{\sigma}} &= \frac{201}{17.14418} \\ &= 11.72409\end{aligned}$$

习题 4.1.2

根据题目中的数据, 有 7 个人体重大于 215, 设 X 是一个对是否大于 215 磅的 0-1 变量, 根据例 4.1.2

$$\begin{aligned}\hat{p} &= \frac{x}{n} = \frac{7}{26} \\ &= 0.26923\end{aligned}$$

若假设模型是正态的, 那么真参数

$$p = P(X > 215)$$

我们前面已经计算出正态分布参数的极大似然估计, 于是

$$\begin{aligned}\hat{p} &= P(X > 215) \\ &= 1 - P\left(\frac{X-\mu}{\sqrt{\sigma^2}} \leq \frac{215-201}{\sqrt{293.9231}}\right) \\ &= 1 - P(Z \leq 0.8166)\end{aligned}$$

```
pnorm(215, mean = mean(x), sd = sqrt(var(x)*25/26), lower.tail = F)
# [1] 0.2070776
```

习题 4.1.6

$$
\begin{aligned}
E\left[\hat{p}\left(a_j\right)\right] &= \frac{1}{n}\sum_{i=1}^{n} E\left[I_j\left(X_i\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} p\left(a_j\right) \\
&= p\left(a_j\right)
\end{aligned}
$$

对于方差, 由于样本是独立同分布的, 所以

$$
\operatorname{Var}\left(\sum_{i=1}^{n} I_j\left(X_i\right)\right) = \sum_{i=1}^{n} \operatorname{Var}\left(I_j\left(X_i\right)\right)
$$

对示性函数还有

$$
E\left(\left(I_j\left(X_i\right)\right)^2\right) = E\left(I_j\left(X_i\right)\right)
$$

所以

$$
\begin{aligned}
\operatorname{Var}\left(\hat{p}\left(a_j\right)\right) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} I_j\left(X_i\right)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}\left(I_j\left(X_i\right)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E\left(\left(I_j\left(X_i\right)^2\right) - E\left(I_j\left(X_i\right)\right)^2\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(p\left(a_j\right) - \left(p\left(a_j\right)\right)^2\right) \\
&= \frac{1}{n^2} np\left(a_j\right)\left(1 - p\left(a_j\right)\right) \\
&= \frac{p\left(a_j\right)\left(1 - p\left(a_j\right)\right)}{n}
\end{aligned}
$$

习题 4.1.8

例 4.1.6(模拟泊松变量) 下面 30 个数据是源自泊松分布均值 $\lambda=2$ 的模拟值，参看例 4.8.2 关于泊松变量的生成.

$$\begin{matrix} 2 & 1 & 1 & 1 & 1 & 5 & 1 & 1 & 3 & 0 & 2 & 1 & 1 & 3 & 4 \\ 2 & 1 & 2 & 2 & 6 & 5 & 2 & 3 & 2 & 4 & 1 & 3 & 1 & 3 & 0 \end{matrix}$$

其 pmf 的非参数估计值是

j	0	1	2	3	4	5	$\geqslant 6$
$\hat{p}(j)$	0.067	0.367	0.233	0.167	0.067	0.067	0.033

习题 4.1.8

先求参数 λ 的极大似然估计，

$$PMF : p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \ldots.$$

于是似然函数，

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} p(x_i;\lambda) \\ &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\ &= \frac{e^{-n\lambda}\lambda^{x_1+x_2+\ldots+x_n}}{x_1!x_2!\ldots x_n!} \end{aligned}$$

对数似然

$$
\begin{aligned}
l(\lambda) &= \ln(L(\lambda)) \\
&= \ln\left(\frac{e^{-n\lambda}\lambda^{x_1+x_2+\ldots+x_n}}{x_1!x_2!\ldots x_n!}\right) \\
&= \ln\left(e^{-n\lambda}\lambda^{x_1+x_2+\ldots+x_n}\right) - \ln\left(x_1!x_2!\ldots x_n!\right) \\
&= \ln\left(e^{-n\lambda}\right) + \ln\left(\lambda^{x_1+x_2+\ldots+x_n}\right) - \ln\left(x_1!x_2!\ldots x_n!\right) \\
&= -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)
\end{aligned}
$$

求导

$$
\frac{\partial l(\lambda)}{\partial\lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0
$$

$$
\Rightarrow \lambda = \bar{x}
$$

在本题中

$$\hat{\lambda} = \bar{x} = \frac{1}{30}(2+1+1+\ldots+1+3+0) = 2.133$$

$p(0;\hat{\lambda})$	$= \frac{2.133^0 e^{-2.133}}{0!}$		$= 0.118$
$p(1;\hat{\lambda})$	$= \frac{2.133^1 e^{-2.133}}{1!}$		$= 0.253$
$p(2;\hat{\lambda})$	$= \frac{2.133^2 e^{-2.133}}{2!}$		$= 0.270$
$p(3;\hat{\lambda})$	$= \frac{2.133^3 e^{-2.133}}{3!}$		$= 0.192$
$p(4;\hat{\lambda})$	$= \frac{2.133^4 e^{-2.133}}{4!}$		$= 0.102$
$p(5;\hat{\lambda})$	$= \frac{2.133^5 e^{-2.133}}{5!}$		$= 0.044$

习题 4.1.8

比较参数模型和非参数模型给出的结果:

x	0	1	2	3	4	5	≥ 6
$\hat{p}(x)$	0.067	0.367	0.233	0.167	0.067	0.067	0.033
$p(x;\hat{\lambda})$	0.118	0.253	0.270	0.192	0.102	0.044	0.022

最右下角的一块就是极大似然估计.

问题

你认为那种是 "对" 的?

习题 4.2.4

$\Gamma(1,\theta)$ 的矩母函数是

$$m(t) = E\left[e^{tX}\right] = (1-\theta t)^{-1}$$

从而 $\dfrac{2X}{\theta}$ 的矩母函数是

$$E\left[e^{(2t/\theta)X}\right] = m(2t/\theta) = (1-2t)^{-1}$$

服从 $\chi^2(2)$ 分布, 因此

$$\sum_{i=1}^{n} 2X_i/\theta \sim \chi^2(2n)$$

习题 4.2.4

用上面的随机变量做枢轴量, 于是一个自然的构造是

$$
\begin{aligned}
& w_{1-(\alpha/2),2n} \leq \frac{2}{\theta}\sum_{i=1}^{n} X_i \leq w_{\alpha/2,2n} \\
\Leftrightarrow\ & \frac{2}{w_{\alpha/2,2n}}\sum_{i=1}^{n} X_i \leq \theta \leq \frac{2}{w_{1-(\alpha/2),2n}}\sum_{i=1}^{n} X_i
\end{aligned}
$$

其中

$$
\begin{aligned}
P\left(w_{1-(\alpha/2),2n} \leq W \leq w_{\alpha/2,2n}\right) &= 1-(\alpha/2)-(\alpha/2) \\
&= 1-\alpha
\end{aligned}
$$

W 服从 $\chi^2(2n)$ 分布.

习题 4.2.4

4.1.1 将 20 个马达放置于高温环境下进行试验. 在这些条件下，马达的寿命用小时表示，由下面数据给出. 假如我们认为，在这些条件下马达寿命 X 服从 $\Gamma(1, \theta)$分布.

1	4	5	21	22	28	40	42	51	53
58	67	95	124	124	160	202	260	303	363

$$\sum X_i = 2021, \quad n = 20$$

```
x<-c(1,4,5,21,22,28,40,40,51,53,58,67,95,124,124,160,202,260,303,363)
> 2 / qchisq(0.025, 40) * sum(x)
[1] 165.4317
> 2 / qchisq(0.975, 40) * sum(x)
[1] 68.11398
```

习题 4.2.4

对比大样本的置信区间:

$$
\begin{aligned}
95\% C.I &= \bar{x} \pm t_{\text{critical}} \left(\frac{s}{\sqrt{n}} \right) \\
&= 101.15 \pm 2.093 \left(\frac{105.4}{\sqrt{20}} \right) \\
&= 101.15 \pm 49.33 \\
&= (51.82, 150.48)
\end{aligned}
$$

问题

你会做怎样的选择?

习题 4.2.10

(a)

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{9}} \sim N(0,1)$$

于是

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{9}}$$

(b)

$$\frac{\bar{X}-\mu}{S/\sqrt{9}} \sim t(8)$$

于是

$$\bar{X} \pm t_{\alpha/2}(8)\frac{S}{\sqrt{9}}$$

提示是在说怎么把 S 写成和 χ^2 有关, 把期望去掉.

习题 4.2.11

$\bar{X} - X_{n+1} \sim N\left(0, \left(\frac{n+1}{n}\right)\sigma^2\right)$ 从而,

$$\frac{(\bar{X} - X_{n+1})/(\sqrt{\sigma^2(n+1)/n}) \sim N(0,1)}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\Big/(n-1)} \sim \chi^2(n-1), \perp\!\!\!\perp} \sim t(n-1)$$

因此,

$$k = \sqrt{\frac{n}{n+1}}$$

习题 4.2.17

用大样本构造置信区间 (根据独立同分布的大数定律), $X_i \sim_{iid} Poi(\mu)$, 根据 $\hat{\mu} = \bar{X}$, 于是 $\mathbb{E}\hat{\mu} = \mu$, $Var(\hat{\mu}) = \frac{1}{n}\mu$, 从而 $\widehat{Var}(\hat{\mu}) = \frac{\bar{X}}{n}$, 因此, **一个近似的**$(1-\alpha)\%$ 置信区间是

$$CI = \left(\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}}{n}}, \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}}{n}}\right)$$

因此

$$\begin{aligned} CI &= \left(\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}}{n}}, \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}}{n}}\right) \\ &= \left(3.4 - (1.645)\sqrt{\frac{3.4}{200}}, 3.4 + (1.645)\sqrt{\frac{3.4}{200}}\right) \\ &= (3.4 - 0.21, 3.4 + 0.21) \\ &= (3.19, 3.61) \end{aligned}$$

方差的置信区间.

$$\chi^2_{(1-\frac{a}{2})}(n-1) \le \frac{(n-1)s^2}{\sigma^2}$$

$$\Rightarrow \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2})}(n-1)}$$

$$\frac{(n-1)s^2}{\sigma^2} \le \chi^2_{(\frac{\alpha}{2})}(n-1)$$

$$\Rightarrow \sigma^2 \ge \frac{(n-1)s^2}{\chi^2_{(\frac{\alpha}{2})}(n-1)}$$

$$1-\alpha = P\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

$$
\begin{aligned}
95\%C.I &= \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \\
&= \frac{(9-1)7.93}{17.53} \le \sigma^2 \le \frac{(9-1)7.93}{2.18} \\
&= \left(3.625 \le \sigma^2 \le 29.101\right)
\end{aligned}
$$

4.2.18

当 μ 已知,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

服从标准正态分布, 于是

$$1-\alpha = P\left(-z_{\alpha/2} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right)$$

逆转即可.

习题 4.2.19

Γ 分布置信区间. 考察 $\frac{2X}{\beta}$ 的矩母函数

$$\begin{aligned} M(t) &= E\left(\exp\left(t\frac{2X}{\beta}\right)\right) \\ &= M_X\left(\frac{2t}{\beta}\right) \\ &= \left[1-\beta\left(\frac{2t}{\beta}\right)\right]^{-3} \\ &= (1-2t)^{-3} \end{aligned}$$

从而, $\frac{2X}{\beta} \sim \chi^2(6)$, $\sum_{i=1}^{n} \frac{2X_i}{\beta} \sim \chi^2(6n)$.

习题 4.2.19

$$P\left(a < \frac{2}{\beta}\sum_{i=1}^{n} X_i < b\right) = P\left(\frac{2}{b}\sum_{i=1}^{n} X_i < \beta < \frac{2}{a}\sum_{i=1}^{n} X_i\right)$$
$$= 0.95$$

置信区间

$$\left(\frac{2}{b}\sum_{i=1}^{n} X_i, \frac{2}{a}\sum_{i=1}^{n} X_i\right)$$

方差已知的均值之差的置信区间.

$$\mathbb{E}(\bar{X}-\bar{Y})=E(\bar{X})-E(\bar{Y})=\mu_1-\mu_2$$

$$\begin{aligned}Var(\bar{X}-\bar{Y}) &= Var(\bar{X})+Var(\bar{Y})-2\,\mathrm{Cov}(\bar{X},\bar{Y})\\ &= Var(\bar{X})+Var(\bar{Y})(\mathrm{Cov}(\bar{X},\bar{Y})=0,\ \text{由于两样本之间的独立性})\\ &= \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\end{aligned}$$

根据独立同分布的大数定律

$$(\bar{X}-\bar{Y})\overset{D}{\sim}N\left((\mu_1-\mu_2),\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

习题 4.2.23

于是根据大样本理论,

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \overset{D}{\sim} N(0, 1)$$

为使

$$P\left(-Z_{\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < Z_{\alpha/2} \right) = 1 - \alpha$$

即

$$\mathrm{P}\left((\bar{X} - \bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X} - \bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 1 - \alpha$$

即可得出 $(1 - \alpha)\%$ 置信区间.

习题 4.2.25

$X_i \sim_{iid} N(\mu_1, \sigma^2)$, $Y_l \sim_{iid} N(\mu_2, \frac{\sigma^2}{3})$. 由于 X_i, Y_j 均是正态分布, 因此

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{9}\right) \quad \bar{Y} \sim N\left(\mu_2, \frac{\sigma^2}{36}\right)$$

根据 4.2.13 式

$$\left((\bar{x} - \bar{y}) - t_{\alpha/2, n-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{x} - \bar{y}) + t_{\alpha/2, n-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

带入 n_1, n_2, S_p 即可.

习题 4.2.27

$X_1, \cdots, X_n \sim N(\mu_1, \sigma_1^2)$, $Y_1, \cdots, Y_m \sim N(\mu_2, \sigma_2^2)$.

$$F = \frac{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)}{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)} = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \sim F(m-1, n-1)$$

由于要求

$$P(F < b) = 0.975, P(a < F < b) = P(F < b) - P(F < a)$$

因此可以取

$$a = F_{0.025}(m-1, n-1), b = F_{0.975}(m-1, n-1)$$

得到 95% 置信区间.

$$\left(F_{0.025}(m-1,n-1)\frac{S_1^2}{S_2^2}, F_{0.975}(m-1,n-1)\frac{S_1^2}{S_2^2}\right)$$