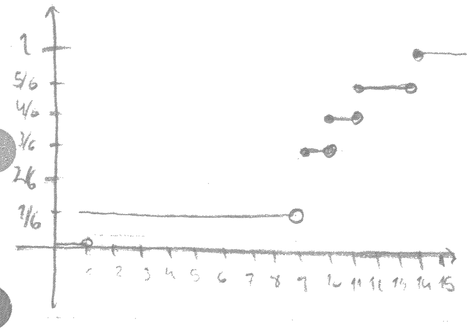
Chapter 10

① Plot the ecdf of 1, 14, 10, 9, 11, 9

$F_n(x) = \frac{1}{n} (\# X_i \leq x)$

So, for $X_{(1)}, X_{(2)}, \ldots, X_{(n)} = 1, 9, 9, 10, 11, 14$




③ Upper and lower quartiles and the median of the distribution of the melting points?

$q_{0.25} \approx 63.4 \quad q_{0.5} \approx 63.6 \quad q_{0.75} \approx 63.8$

⑨ Use the method of propogation of error to derive an approximation to the bias of the log $\overset{\text{empirical}}{\phantom{x}}$ survival function. Where is the bias large and what is its sign?

The survival function is $S(x) = \mathbb{P}(X > x) = 1 - F(x)$

Consider now a random variable $T$ and a function $g$ and define $Y = g(T)$

We can approximate $g$ through a Taylor series expansion (method of propogation) about $\mu_T$ as

$Y = g(T) \approx g(\mu_T) + (X - \mu_T) g'(\mu_T) + \frac{1}{2}(X - \mu_T)^2 g''(\mu_T)$

Taking expectation on both sides we get

$E(Y) \approx g(\mu_T) + \frac{\sigma_T^2}{2} g''(\mu_T) \Rightarrow E(Y) - g(\mu_T) \approx \frac{1}{2}\sigma_T^2 g''(\mu_T) \Rightarrow$

Bias $\approx \frac{1}{2}\sigma_T^2 g''(\mu_T)$

In this case: $Y = \log(1 - F_n(x))$ $T = 1 - F_n(x)$ $g(t) = \log(t)$

$\sigma_T^2 = \mathrm{Var}(1 - F_n(x)) = \frac{1}{n} F(x)[1 - F(x)]$ *

$g''(t) = -\frac{1}{t^2} \quad \mu_T = 1 - F(x)$

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$ $*$

$$I_{(-\infty,x]}(X_i) = \begin{cases} 1 & \text{if } X_i \le x \\ 0 & \text{if } X_i > x \end{cases}$$

$\Rightarrow I_{(-\infty,x]}(X_i) \sim Ber(F(x)) \Rightarrow nF_n(x) \sim Bin(n, F(x))$

$\Rightarrow E[nF_n(x)] = nF(x)$ and $Var(nF_n(x)) = nF(x)(1-F(x))$

$E[F_n(X)] = F(X) \Rightarrow$

$\mu_T = E[1-F_n(X)] = 1 - F(X)$

$$Var[F_n(X)] = \frac{F(X)[1 F(X)]}{n} \Rightarrow$$

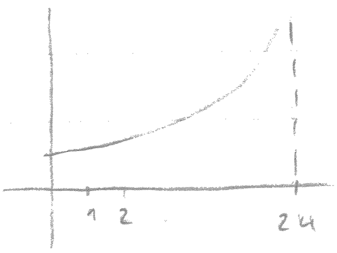
$$Var[1-F_n(X)] = \frac{1}{n} F(x)[1-F(x)]$$

$$Bias \approx \frac{1}{2n} F(x)[1-F(x)]\left(\frac{-1}{[1-F(x)]^2}\right)$$

$$= \frac{-1}{2n}\frac{F(X)}{1-F(X)}$$ which is negative and large for large $X$

(15) $T \sim U[0,24] \Rightarrow f(t) = \frac{1}{24} \Rightarrow F(t) = \frac{t}{24}$ $\quad 0 \le t \le 24$

By definition $h(t) = \dfrac{f(t)}{1-F(t)} = \dfrac{1/24}{1-t/24} = \dfrac{1}{24-t}$




smallest for 0

largest for 24 $\Rightarrow$ probability of release becomes higher the longer he's been waiting

The pth quantile is defined as:

(17)

$F(x) = p$ $\qquad\qquad$ $G(x) = p$

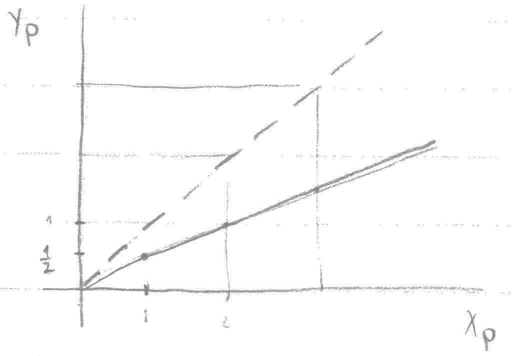
$1 - e^{-t} = p$ $\qquad\qquad$ $1 - e^{-2t} = p$

$-e^{-t} = p - 1$ $\qquad\qquad$ $-e^{-2t} = p - 1$

$e^{-t} = 1 - p$ $\qquad\qquad$ $e^{-2t} = 1 - p$

$-t = \log(1-p)$ $\qquad\qquad$ $-2t = \log(1-p)$

$X_p = -\log(1-p)$ $\qquad\qquad$ $Y_p = -\frac{1}{2}\log(1-p) \Rightarrow Y_p = \frac{1}{2} X_p$




(25) Let $Y_p = c X_p \Rightarrow X_p = \frac{Y_p}{c}$ for all p

$p = F(X_p) = F\left(\frac{Y_p}{c}\right)$ on the other hand $p = G(Y_p) \Rightarrow$

$F\left(\frac{Y_p}{c}\right) = G(Y_p) \Rightarrow F\left(\frac{y}{c}\right) = G(y)$

(29) $X_1, \dots, X_{26}$ 5 are outliers $N$: # of outliers in a bootstrap sample

a) Let $A_i$: the ith observation of the bootstrap sample is an outlier

then $A_i \sim \text{Bernoulli}\left(\frac{5}{26}\right)$ $\frac{5}{26}$ probability of success

then $N = \sum_{i=1}^{26} I(A_i)$ where $I(A_i)$ indicator function of $A_i$ $\Rightarrow$

$N \sim \text{Bin}\left(26, \frac{5}{26}\right)$

b) $P(N \geq 10) = 1 - P(N < 9) = 1 - \sum_{i=1}^{9} \binom{26}{i} \frac{5}{26}^{i} \left(1 - \frac{5}{26}\right)^{26-i} = 1 - 0.9821 = 0.0179$

c) Since each of the 1000 bootstrap samples is constructed independently, the number of them that contain 10 or more outliers follows a binomial distribution of 1000 trials an probability 0.0179 of success. That distribution has expected value $np = 1000 \cdot 0.0179 \approx 18$

d) $P(N=26) = \binom{26}{20}\left(\frac{5}{20}\right)^{26}\left(1-\frac{5}{20}\right)^{26-26} = \left(\frac{5}{26}\right)^{26} = 2.42 \times 10^{-19}$

(33) Mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{10} X_i$ so it considers all the numbers in the sample ✓

Median middle value of ordered observations $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ X

10% trimmed mean: $\bar{X}_\alpha = \dfrac{X_{([n\alpha]+1)} + \ldots + X_{(n-[n\alpha])}}{n-2[n\alpha]}$

$\bar{x}_{0.1} = \dfrac{X_{(11)} + X_{(12)} + \ldots + X_{(90)}}{80}$ X

Std: $s = \sqrt{\dfrac{1}{n-1}\sum(X_i - \bar{X})^2}$ ✓

Median absolute deviation from the median

**MAD: median of $|X_i - \tilde{X}|$** X

IQR: $q_{0.75} - q_{0.25}$ X

(41) Let $x_p$ be the $p^{th}$ quantile, then $P(X < x_p) = p$ and $P(X \geq x_p) = 1-p$

Let $X_{(s)}$ be the sth order statistic of $X_1, X_2, \ldots, X_n$, then

$P(X_{(s)} \geq x_p) = 1 - P(X_{(s)} < x_p) = 1 - P(\text{at least } s \text{ observations} < x_p)$

$= 1 - \sum_{i=s}^{n}\binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^{s-1}\binom{n}{i} p^i (1-p)^{n-i}$

Similarly $P(X_{(r)} \leq x_p) = P(\text{at least } r \text{ observations} < x_p)$

$= \sum_{i=r}^{n}\binom{n}{i} p^i (1-p)^{n-i}$ Therefore $P(X_{(r)} \leq x_p \leq X_{(s)}) = \mathbb{P}(X_{(r)} < x_p \text{ and } X_{(s)} \geq x_p)$

$= 1 - \mathbb{P}(X_{(r)} > x_p \text{ or } X_{(s)} < x_p) = 1 - [\mathbb{P}(X_{(r)} > x_p) + \mathbb{P}(X_{(s)} < x_p)] = 1 - [1 - \mathbb{P}(X_{(r)} \leq x_p) + 1 - \mathbb{P}(X_{(s)} \geq x_p)]$

$= 1 - 1 + \mathbb{P}(X_{(r)} \leq x_p) - 1 + \mathbb{P}(X_{(s)} \geq x_p) = \sum_{i=r}^{n}\binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=0}^{s-1}\binom{n}{i} p^i (1-p)^{n-i} - 1 = \sum_{i=r}^{s-1}\binom{n}{i} p^i (1-p)^{n-i}$

So $(X_{(r)}, X_{(s)})$ is a CI for $x_p$ with $\sum_{i=r}^{s-1}\binom{n}{i} p^i (1-p)^{n-i}$ probability of coverage

an $100(1-\alpha)\%$ confidence interval requires r and s to be taken such that

$\sum_{i=r}^{s-1} p^i (1-p)^{n-i} \geq 1-\alpha$

4.11 Let $x_p$ be the $p$-th quantile of a sample $X_1, X_2, \dots, X_n$, then $\mathbb{P}(X_i < x_p) = p$. Let $X_{(r)}$ and $X_{(s)}$ be the $r$th and $s$th order statistics of the sample and $r < s$. Then

$$\mathbb{P}(X_{(s)} < x_p) = \mathbb{P}(\text{at least } s \text{ observations} < x_p) = \sum_{i=s}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Similarly

$$\mathbb{P}(X_{(r)} > x_p) = 1 - \mathbb{P}(X_{(r)} \leq x_p) = 1 - \mathbb{P}(\text{at least } r \text{ observations} \leq x_p)$$

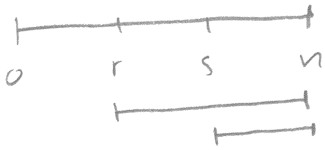
$$= 1 - \sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Then

$$\mathbb{P}(X_{(r)} \leq x_p \leq X_{(s)}) = \mathbb{P}(X_{(r)} \leq x_p \text{ and } X_{(s)} \geq x_p) = 1 - \mathbb{P}(X_{(r)} > x_p \text{ or } X_{(s)} < x_p)$$

$$= 1 - \left[\mathbb{P}(X_{(r)} > x_p) + \mathbb{P}(X_{(s)} < x_p)\right] = 1 - \left[1 - \sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=s}^{n} \binom{n}{i} p^i (1-p)^{n-i}\right]$$

$$= \sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i=s}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$




$$= \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}$$

So $(X_{(r)}, X_{(s)})$ is a confidence interval for $x_p$ with $\sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}$ probability of coverage. An $100(1-\alpha)\%$ confidence interval requires thus $r$ and $s$ to be taken such that $\sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} = 1-\alpha$

Chapter 11

(7) Let $m=n$ and consider the first case, then

$$Var(\bar{X}-\bar{Y}) = S_p^2\left(\frac{1}{n}+\frac{1}{n}\right) = S_p^2 \frac{2}{n} \quad \text{where } S_p^2 = \frac{(n-1)S_x^2+(m-1)S_y^2}{m+n-2} = \frac{(n-1)(S_x^2+S_y^2)}{2(n-1)}$$

$$= \frac{S_x^2+S_y^2}{2}$$

then $V(\bar{X}-\bar{Y}) = \frac{S_x^2+S_y^2}{2}\cdot\frac{2}{n} = \frac{S_x^2+S_y^2}{n}$

For the second case

$Var(\bar{X}-\bar{Y}) = \frac{S_x^2}{n} + \frac{S_y^2}{n} = \frac{S_x^2+S_y^2}{n}$ $\therefore$ the estimates are identical

(13) $X_1,\dots,X_{25} \sim iid\ N(0.3,1)$ Test $H_0: \mu=0$ at significance level $\alpha=0.05$
$H_A: \mu>0$

The power of the test is the probability of rejecting the null hypothesis when it is false. In this case, the null hypothesis is rejected for large values of the test statistic $t = \frac{\bar{X}}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}}{\frac{1}{5}} = 5\bar{X}$ We reject $H_0$ if $t \geq z(\alpha)$, that is

$P(t \geq z_{0.95}) = P(5\bar{X} \geq z_{0.95}) = P(\bar{X} \geq \frac{1}{5} z_{0.95}) =$

$P\left(\frac{\bar{X}-0.3}{\sigma_{\bar{X}}} \geq \frac{\frac{1}{5}z_{0.95}-0.3}{\sigma_{\bar{X}}}\right) = P\left(Z \geq \frac{\frac{1}{5}z_{0.95}-0.3}{\frac{1}{5}}\right) = 1-\phi(0.15) = 0.46$

↑assuming $H_0$ is false $\qquad z_{0.95} = 1.65$

Under $H_0$ we would expect to have a balanced number of positive and negative values in the sample. Then the random variable $W$: # of positive signs is $Bin(25, 0.6)$ distributed*. We reject $H_0$ if $W \geq z(\alpha) = z_{0.95} = 16$

$P(W \geq 16) = P\left(\frac{W-25\cdot0.6}{\sqrt{25\cdot0.6(1-0.6)}} \geq \frac{16-25\ 0.6}{\sqrt{25(0.6)(1-0.6)}}\right) = P(Z \geq 0.4) = 1-\phi(0.4) = 0.35$

* $P(X>0) = P(X-0.3>-0.3) = 1-\phi(-0.3) = \phi(0.3) = 0.6$
This is valid under the assumption $H_0$ is false

21) a) $H_0: \mu_I = \mu_{II}$ $\quad \bar{X} = 10.6930 \quad S_x^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2 = 23.2255$

$H_1: \mu_I \neq \mu_{II}$ $\quad \bar{Y} = 6.7500 \quad S_y^2 = 12.4775$

Construct $t = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{S_x^2}{n} + \dfrac{S_y^2}{m}}} = \dfrac{10.6930 - 6.75}{\sqrt{\dfrac{23.2255}{10} + \dfrac{12.4775}{10}}} = 2.0723$

At 95% confidence

$t_{18}(0.025) = 2.1 \Rightarrow$ don't reject

b)

| Type I | Type II |
|---|---|
| 3.03 (1) | 3.19 (2) |
| 5.53 (8) | 4.26 (3) |
| 5.60 (9) | 4.47 (4) |
| 9.30 (11) | 4.53 (5) |
| 9.92 (13) | 4.67 (6) |
| 12.51 (14) | 4.69 (7) |
| 12.95 (17) | 12.78 (16) |
| 15.21 (18) | 6.79 (10) |
| 16.04 (19) | 9.37 (12) |
| 16.68 (20) | 12.75 (15) |

$R = 1+8+9+11+13+14+17+18+19+20 = 130$

$R' = 20(21)/2 - 130 = 80$

$\Rightarrow R^* = \min(R, R') = 80$

The test rejects for small $R^*$, that is, if $R^* < r(\alpha)$

For $\alpha = 0.05$ $\quad r(\alpha) = 78$

but $R^* > 78$ so the test doesn't reject

c) Small sample size suggests nonparametric methods are better

d) $\pi = P(X > Y) \Rightarrow \hat{\pi} = \dfrac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} Z_{ij}$ where $Z_{ij} = \begin{cases} 1 & \text{if } X_i > Y_j \\ 0 & \text{otherwise} \end{cases}$

$= \dfrac{1}{10^2}(6+6+7+8+8+10+10+10+10)$

$= 0.75$

(15) The confidence interval for the difference $\bar{X}-\bar{Y}$ is given by

$$(\bar{X}-\bar{Y}) \pm z(\alpha/2)\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}$$

In this case $(\bar{X}-\bar{Y}) \pm z(\alpha/2)\sigma\sqrt{\frac{2}{n}}$

If we want its width to be equal to 2 then

$$2z(\alpha/2)\sigma\sqrt{\frac{2}{n}} = 2 \iff 1.96 \cdot 10 \cdot \sqrt{\frac{2}{n}} = 1 \iff 19.6\sqrt{2} = \sqrt{n} \iff$$

$n = 768.32 \Rightarrow$ take $n = 768$

(17) $H_0: \mu_X = \mu_Y$

$H_A: \mu_X \neq \mu_Y$

We reject if $\left|\frac{(\bar{X}-\bar{Y})}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right| = \left|\frac{(\bar{X}-\bar{Y})}{10\sqrt{\frac{2}{n}}}\right| = \left|\frac{\sqrt{n}(\bar{X}-\bar{Y})}{10\sqrt{2}}\right| > z(\alpha/2) \Rightarrow$

$|\bar{X}-\bar{Y}| > \frac{10\sqrt{2}\,z(\alpha/2)}{\sqrt{n}}$ Take $\delta = \mu_X - \mu_Y$

Power Function $B(\delta) = P(\text{reject} \mid H_A) = P\left[|\bar{X}-\bar{Y}| > \frac{10\sqrt{2}\,z(\alpha/2)}{\sqrt{n}} \,\middle|\, \bar{X}-\bar{Y} \sim N\left(\delta, \frac{10\sqrt{2}}{\sqrt{n}}\right)\right]$

$$= P\left[\frac{\bar{X}-\bar{Y}-\Delta}{10\frac{2}{\sqrt{n}}} > \frac{\frac{10\sqrt{2}}{\sqrt{n}}z(\alpha/2)-\Delta}{10\sqrt{\frac{2}{n}}}\right] + P\left[\frac{\bar{X}-\bar{Y}-\Delta}{10\sqrt{\frac{2}{n}}} < \frac{-\frac{10\sqrt{2}}{\sqrt{n}}z(\alpha/2)-\Delta}{10\sqrt{\frac{2}{n}}}\right]$$

$$= 1 - \Phi\left(z(\alpha/2) - \frac{\sqrt{n}\,\delta}{10\sqrt{2}}\right) - \Phi\left(-z(\alpha/2) - \frac{\sqrt{n}\,\delta}{10\sqrt{2}}\right)$$

a) $\alpha = 0.05$, $n = 20 \Rightarrow z(\alpha/2) = 1.96$

$B_1(\delta) = 1-\Phi\left(1.96 - \frac{\delta}{\sqrt{10}}\right) + \Phi\left(-1.96 - \frac{\delta}{\sqrt{10}}\right)$

b) $\alpha = 0.1$, $n = 20 \Rightarrow z(\alpha/2) = 1.64$

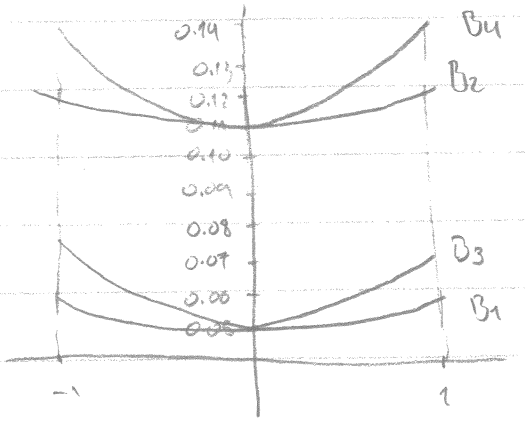
$B_2(\delta) = 1-\Phi\left(1.64 - \frac{\delta}{\sqrt{10}}\right) + \Phi\left(-1.64 - \frac{\delta}{\sqrt{10}}\right)$

c) $\alpha = 0.05$, $n = 40 \Rightarrow z(\alpha/2) = 1.96$

$$B_3(\delta) = 1 - \Phi\left(1.96 - \frac{\delta}{\sqrt{5}}\right) + \Phi\left(-1.96 - \frac{\delta}{\sqrt{5}}\right)$$

d) $\alpha = 0.1$, $n = 40 \Rightarrow z(\alpha/2) = 1.64$

$$B_4(\delta) = 1 - \Phi\left(1.64 - \frac{\delta}{\sqrt{5}}\right) + \Phi\left(-1.64 - \frac{\delta}{\sqrt{5}}\right)$$




- increase $n \Rightarrow$ uniformely more powerful test
- increase $\alpha \Rightarrow$ uniformely more powerful test