6.4. Distribution - free methods.

I. Misspecification of the family $\{P_\theta : \theta \in \Omega\}$ is far from the truth.

II. Inferences about discrete families are easier: if $\theta \in \Omega = [0,1]$ in the Bernoulli model, can't misspecify because the model includes ALL distr. on $S = \{0,1\}$.
The same with other discrete distr.

6.4.1 Method of moments.

Take $\{P_\theta : \theta \in \Omega\}$ the set of all distr with l finite moments, want to make inferences about $\mu_i = \mathbb{E}_\theta(X^i)$, for $i = 1, \dots, l$, based on a sample $(X_1, \dots, X_n)$. The natural sample analog is the ith sample moment

$$m_i = \frac{1}{n}\sum_{j=1}^{n} X_j^i \text{ ; note } \mathbb{E}_\theta(M_i) = \mu_i \quad \forall \theta \in \Omega,$$

WLLN & SLLN are true. Moreover,

$$\frac{M_i - \mu_i}{\sqrt{Var_\theta(M_i)}} \xrightarrow{D} N(0,1), \text{ as } n \to \infty \text{ if } Var_\theta(M_i) < \infty$$

Check that $Var_\theta(M_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var_\theta X_j^i = \frac{1}{n} Var_\theta (X_1^i) =$

$$= \frac{1}{n}\mathbb{E}_\theta\left((X_1^i - \mu_i)^2\right) = \frac{1}{n}\mathbb{E}_\theta\left(X_1^{2i} - 2\mu_i \cdot X_1^i + \mu_i^2\right) =$$

$$= \frac{1}{n}\left(\mu_{2i} - \mu_i^2\right) < \infty \text{ if } i \le \frac{l}{2}.$$

Idea: estimate $\mu_{2i} - \mu_i^2$ by $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j^i - M_i\right)^2$

.. because we think that $X_1^i, \dots, X_n^i$ is a sample from the distribution with mean μ_i and variance $\mu_{2i} - \mu_i^2$

Remark: $\mathbb{E}S_i^2 = \mu_{2i} - \mu_i^2$

$\Rightarrow m_i \pm z_{\frac{1+\delta}{2}} \frac{s_i}{\sqrt{n}}$ is an approx. of CI for large n.

How to use it for estimation of other parameters?!

<u>Ex. 1</u> . $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$;

$\mu = EX$; $\sigma^2 = \text{Var}X$;

, so $\mu = \bar{Y}$

$\mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} Y_i^2$

$\Rightarrow \hat{\mu}_{MME} = \bar{Y}$, $\hat{\sigma}^2_{MME} = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 =$

$= \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$

<u>Ex. 2</u> Uniform distribution $U[0, \theta]$

Let $X_1, \ldots, X_n \sim U[0, \theta]$;

Recall that the MLE estimator is $\hat{\theta} = Y_{max}$; now, $\mu_1 = \frac{\theta}{2}$, $m_1 = \bar{Y}$

$\Rightarrow \frac{\theta}{2} = \bar{Y} \Rightarrow \hat{\theta}_{MME} = 2\bar{Y} \neq Y_{max}$

Note that MME est-ors don't always make sense; e.g. if in Ex. 2,

$x_1 = 0.2$, $x_2 = 0.6$, $x_3 = 0.4$, $x_4 = 2$

from $U(0, \theta)$; here $n = 4$.

Then, $\hat{\theta}_{MME} = 2\bar{x} = 1.6$ but there is an obs.

$x_4 = 2 > \hat{\theta}_{MME}$.

Bootstrap Let $X_1, \ldots, X_n \sim P_\theta$, θ unknown;

$\{P_\theta : \theta \in \Omega\}$

$$\Rightarrow \hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty, x)}(X_i)$$ — a natural estimator of the cdf $F(x)$

$= \cancel{\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \le x)} =$

$$= \frac{1}{n}\#\{\# i : X_i \le x\}$$

$$\Rightarrow \mathbb{E}_\theta \hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} P(X_i \le x) = \frac{1}{n}\cdot nF(x) = F(x)$$

, so $\hat{F}$ is unbiased for F. ; By WLLN we have

$$\hat{F}_\theta(x) \to F(x) \text{ as } n \to \infty$$

Since $\mathbb{I}_{(-\infty, x]}(X_i)$ is a sample from Bernoulli $(F_\theta(x))$ distr, we have that the SD of $\hat{F}(x)$ is $\sqrt{\frac{\hat{F}(x)(1-\hat{F}(x))}{n}}$. Also, observe that $\hat{F}(x)$ prescribes a distribution on $\{X_1, \ldots, X_n\}$ where $\frac{1}{n}$ is a mass at each X_i if they are distinct. If they are not all distinct we have a mass of $\frac{f_i}{n}$ for X_i where f_i is the number of times X_i occurs in the sample $X_1, \ldots, X_n$.

Let us we're interested in estimating some characteristic of the distribution F_θ, e.g. its jth moment, its pth quantile etc. Denote this characteristic $\Psi(\theta) = T(F_\theta)$. Say, $\hat{\Psi}(X_1, \ldots, X_n)$ is an estimator of $\Psi(\theta)$. Accuracy of $\hat{\Psi}$ is important... so we measure it by $MSE_\theta(\hat{\Psi}) = [\mathbb{E}_\theta(\hat{\Psi}) - \Psi(\theta)]^2 + Var_\theta(\hat{\Psi})$ — has to be estimated because it depends on the unknown parameter θ. For example, $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$; $\hat{\mu} = \bar{X}$; $Var\bar{X} = MSE(\bar{X}) = \frac{\sigma^2}{n}$ but σ^2 is unknown... and has to be estimated...

In general, bias estimation is easier (if needed): squared bias is estimated as $[\hat{\varphi} - T(\hat{F})]^2$ (*) because $\hat{F}$ is close to the true F when n is large.

To estimate the variance of $\hat{\varphi}$, we use:

$$Var_{\hat{F}}(\hat{\varphi}) = \mathbb{E}_{\hat{F}}(\hat{\varphi})^2 - \mathbb{E}^2_{\hat{F}}(\hat{\varphi}) = \frac{1}{n^n}\sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} \hat{\varphi}^2(x_{i_1}, \ldots, x_{i_n}) - \left(\frac{1}{n^n}\sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} \hat{\varphi}(x_{i_1}, \ldots, x_{i_n})\right)^2,$$

t.i. $x_1, \ldots, x_n$ are iid random values with the cdf given by $\hat{F}$

Solution: draw m indep. samples of size n from $\hat{F}$, evaluate $\hat{\varphi}$ for each of these samples to obtain $\hat{\varphi}_1, \ldots, \hat{\varphi}_m$; then use the sample variance (**) $\widehat{Var}_{\hat{F}}(\hat{\varphi}) = \frac{1}{m-1}\left\{\sum_{i=1}^{m} \hat{\varphi}_i^2 - \left(\frac{1}{m}\sum_{i=1}^{m} \hat{\varphi}_i\right)^2\right\}$ as the estimate. These are bootstrap samples (resamples). Furthermore, combining (*) & (**) gives the estimated MSE. $\frac{1}{m}\sum_{i=1}^{m}\hat{\varphi}_i$ is the bootstrap mean & $\sqrt{\widehat{Var}_{\hat{F}}(\hat{\varphi})}$ is the bootstrap standard error. The bootstrap standard error is a valid estimate of the error in $\hat{\varphi}$ when $\hat{\varphi}$ has only a limited bias.

Ex 6.4.2 Sampling from a symmetric distribution, we can use the sample median as a robust estimator of the sampling mean – $\hat{x}_{0.5}$ – to estimate the unknown μ. Sampling distribution for $\hat{x}_{0.5}$ is typically hard to obtain. How to estimate MSE?! First squared bias is estimated (recall that $T(F)$ is the mean) by

$$\left(\hat{\varphi} - T(\hat{F})\right)^2 = \left(\hat{x}_{0.5} - \bar{x}\right)^2.$$

This should be close to zero due to symmetry; next, to estimate the variance, we generate m samples of size n from $\{x_1, \ldots, x_n\}$ (with replacement) and calculate $\hat{x}_{0.5}$ for each sample. Note that, for a given sample size $n = 15$, $\bar{x} = -2.087$, the estimated squared bias is, since $\hat{\varphi} = -2.000$, $(-2.000 + 2.087)^2 = 7.569 \cdot 10^{-3}$ – appropriately small.

choose $m = 10^3$, $m = 10^4$, $m = 10^5$

...p Confidence Intervals.

...od 1. $\hat{\psi} \pm t_{\frac{1+\gamma}{2}}(n-1)\sqrt{\widehat{Var}_{\hat{F}}(\hat{\psi})}$

Method 2 Bootstrap percentile CI,

$\left(\hat{\psi}_{(1-\gamma)/2}, \hat{\psi}_{1+\gamma/2}\right)$ – $\hat{\psi}_p$ is the pth empirical percentile

Both need at least approximate unbiasedness of $\hat{\psi}$ for $\psi(\theta)$.

Ex. 6.4.3. 0.25 trimmed mean as an estimator of the population mean.

Let $\lfloor x \rfloor$ – greatest integer $\le x \in \mathbb{R}^1$.

Def For $\alpha \in [0,1)$, a sample α trimmed mean is

$$\bar{X}_\alpha = \frac{1}{n-2\lfloor \alpha n \rfloor} \sum_{i=\lfloor \alpha n \rfloor + 2}^{n-\lfloor \alpha n \rfloor} X_{(i)}$$

– toss out approx. $\lfloor \alpha n \rfloor$ smallest values and $\lfloor \alpha n \rfloor$ of the largest values.

Special cases: $\alpha = 0$; $\alpha = 0.5$.

For the same data as in the previous example, take $\alpha = 0.25$: $(0.25) * 15 = 3.75$, discard 3 obs. on both sides.

The average is, then, $\hat{\psi} = \bar{x}_{0.25} = -1.97778$ ...

Again, perform bootstrap with $m = 10^4$ samples.

Find, $\sqrt{\widehat{Var}_{\hat{F}}(\hat{\psi})} = 0.7380$ – and a histogram looks very normal-like.

$\Rightarrow$ bootstrap t 0.95 CI is $-1.97778 \pm (2.14479)(0.7380) \simeq (-3.6, -0.4)$

$\rightarrow$ Read the vector in, then remove 6 elements by

Xt<-X[-c(1:3, 13:15)]

$\hookrightarrow$ But first X<-sort(X)

6.4.3. The Sign Statistic.

Sample $\{x_1, \dots, x_n\}$. For simplicity only consider $p = \frac{i}{n}$ for some $i \in \{1, \dots, n\}$. This implies that $\hat{x}_p = x_{(i)}$ is the natural estimate of x_p. Let $H_0 : x_p = x_0$. Use the sign test statistic $S = \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x_0]}(x_i)$. S is the # of sample values $\le x_0$. If H_0 is true, $\mathbb{1}_{(-\infty, x_0]}(x_1), \dots, \mathbb{1}_{(-\infty, x_0]}(x_n)$ is a sample from $Ber(p)$ distribution.

Thus, when H_0 is true, $S \sim Bin(n,p)$. This gives us an opportunity to construct a test of H_0; since binomial distribution is unimodal, the P-value will be the prob. of the set $\left\{ i : \binom{n}{i} p^i (1-p)^{n-i} \le \binom{n}{S_0} p^{S_0} (1-p)^{n-S_0} \right\}$

Note that this p-value is independent of any possible parameters of the distribution.

When n is large, under H_0, $Z = \dfrac{S - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1)$

Thus, an approx. P-value is given by

$$2\left\{1 - \Phi\left(\left|\frac{S_0 - 0.5 - np}{\sqrt{np(1-p)}}\right|\right)\right\} \to \text{continuity correction}$$

A special case $p = \frac{1}{2}$ - corresponds to $x_{0.5}$. The distribution of S under H_0 is $Bin\left(n, \frac{1}{2}\right)$ - a unimodal and symmetrical about $\frac{1}{2}$ $\Rightarrow$ The P-value is defined from $\left\{ i : \left|S_0 - \frac{n}{2}\right| \le \left|i - \frac{n}{2}\right| \right\}$.

Now, let j be the smallest integer $\ge \frac{n}{2}$ such that

$$P\left(\left\{ i : \left|i - \frac{n}{2}\right| \ge j - \frac{n}{2} \right\}\right) \le 1 - \delta \text{ where } P \sim Bin\left(n, \tfrac{1}{2}\right).$$

If the observed $S \in \left\{ i : \left|i - \frac{n}{2}\right| \ge j - \frac{n}{2} \right\}$, we reject $H_0 : x_{0.5} = x_0$ at the $1-\delta$ level and will not otherwise.

The corresponding γ - confidence intervals, equivalent to the set of all values $x_{0.5}$ such that $H_0 : x_{0.5}(\theta) = x_0$ is not rejected at the $1-\gamma$ level, is equal to

$$C(x_1, \ldots x_n) = \left\{ x_0 : \left| \sum_{i=1}^{n} I_{(-\infty, x_0]}(x_i) - \frac{n}{2} \right| < j - \frac{n}{2} \right\} =$$

$$= \left\{ x_0 : n-j < \sum_{i=1}^{n} I_{(-\infty, x_0]}(x_i) < j \right\} = [x_{(n-j+1)}, x_{(j)}]$$

because, for example, $n - j < \sum_{i=1}^{n} I_{(-\infty, x_0]}(x_i)$ iff $x_0 \geq x_{(n-j+1)}$.

Ex. 6.4.4. Application of the sign test

Sample size $n=10$ from a continuous distr.

$H_0 : x_{0.5}(\theta) = 0$, the sample is 0.44 −0.06 0.43 −0.16 −2.13 1.15 1.08 5.67 −4.97 0.11

Boxplot indicates 2 extreme obs, so the use of median to estimate its location is prob. warranted. The sample median is

$$\frac{0.11+0.43}{2} = 0.27; \quad S = \sum_{i=1}^{10} I_{(-\infty, 0]}(x_i) = 4$$

The P-value is given by $P(\{i : |4-5| \leq |i-5|\}) =$

$$= P(\{i : |i-5| \geq 1\}) = 1 - P(\{i : |i-5| < 1\}) =$$

$$= 1 - P(\{5\}) = 1 - \binom{10}{5}\left(\frac{1}{2}\right)^{10} = 1 - 0.24609 = 0.75391.$$

To compute the 0.95 CI for the median, first check that for $j=10$, $P(\{i : |i-5| \geq 5\}) = P(\{0, 10\}) = 2\binom{10}{0}\left(\frac{1}{2}\right)^{10} =$
$= 1.9531 \times 10^{-3} < 1 - 0.95 = 0.05$.

For $j=9$, we have

$$P(\{i : |i-5| \geq 4\}) = P(\{0, 1, 9, 10\}) = 2\binom{10}{0}\left(\frac{1}{2}\right)^{10} + 2\binom{10}{1}\left(\frac{1}{2}\right)^{10} = 2.1484 \cdot 10^{-2} < 0.05$$

For $j=8$, check that the result is > 0.05. The appropr. value is $j=9$ and the CI is $[x_{(2)}, x_{(9)}] = [-0.16, 1.15]$