UNIVERSITÉ DE GENÈVE
GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

CICG CENTRE INTERNATIONAL DE CONFÉRENCES GENÈVE

Monte Carlo approach

Idea: For F an unknown distribution function, if we know how to simulate from F, simulate $X_1, \dots, X_B \overset{iid}{\sim} F$ and estimate F by e.c.d.f. $\hat{F}_B(t) = \frac{1}{B} \sum_{i=1}^{B} I[X_i \leq t]$. Glivenko-Cantelli gives

$$\lim_{B \to \infty} \| F - \hat{F}_B \|_\infty = 0 \text{ a.s.}$$

Example 57: Kolmogorov-Smirnov

$X_1, \dots, X_n \sim F$ $\quad H_0: F \sim N(\mu, \sigma^2)$ for $\mu \in \mathbb{R}, \sigma^2 > 0$ known
$\quad H_1: \neg H_0$

$T_n = \| F - \hat{F}_n \|_\infty$ for F_n the e.c.d.f of $X_1, \dots, X_n$

Under H_0, $T_n = \sup_{x \in \mathbb{R}} \left| F(x) - \frac{1}{n} \sum_{i=1}^{n} I[X_i \leq x] \right|$ for $X_1, \dots, X_n \overset{iid}{\sim} F$.

We don't know the exact distribution of T_n, but we can simulate it.

- sample many times $X_1, \dots, X_n \overset{iid}{\sim} F$, each time compute T_n denoted by $T^*_{n,b}$ for b-th run, $b = 1, \dots, B$ $(= 10^4)$
- from all the $\{T^*_{n,b}\}_{b=1}^{B}$ estimate e.c.d.f to approximate the distribution function of T_n under H_0.
- use quantile of this estimated distribution to test.

Example 58: Two sample problem for exponential distribution

$X_1, \dots, X_{n_1} \sim \mathcal{E}xp(\lambda_x)$ iid, independent samples. $\quad H_0: \lambda_x = \lambda_y$
$Y_1, \dots, Y_{n_2} \sim \mathcal{E}xp(\lambda_y)$ $\quad H_1: \lambda_x \neq \lambda_y$

point estimators $\hat{\lambda}_x = 1/\bar{X}$, $\hat{\lambda}_y = 1/\bar{Y}$ $\qquad$ Test statistic $T = \dfrac{\bar{Y}}{\bar{X}} = \dfrac{\hat{\lambda}_x}{\hat{\lambda}_y}$

We have $T = \frac{\bar{Y}}{\bar{X}} \sim \frac{m_1}{m_2} \frac{\sum \Gamma(1, 1/\lambda_y)}{\sum \Gamma(1, 1/\lambda_x)} = \frac{m_1}{m_2} \frac{\Gamma(m_2, 1/\lambda_y)}{\Gamma(m_1, 1/\lambda_x)}$

$$= \frac{m_1/\lambda_y}{m_2/\lambda_x} \frac{\Gamma(m_2, 1)}{\Gamma(m_1, 1)} = \underbrace{\frac{\lambda_x}{\lambda_y}}_{\substack{\text{depends only on} \\ \lambda_x/\lambda_y \text{ for any} \\ \lambda_x, \lambda_y}} \cdot \underbrace{\frac{\Gamma(m_2, m_1)}{\Gamma(m_1, m_2)}}_{\substack{\text{does not depend} \\ \text{on } \lambda_x, \lambda_y \\ (*)}} = \frac{\lambda_x}{\lambda_y} \cdot \frac{\Gamma(\frac{2m_2}{2}, 2)}{\Gamma(\frac{2m_1}{2}, 2)} \cdot \frac{\frac{m_1}{2}}{\frac{m_2}{2}}$$

$$\underset{\Gamma(\frac{\nu}{2}, 2) = \chi^2(\nu)}{=} \frac{\lambda_x}{\lambda_y} \underbrace{\frac{\chi^2(2m_2)/2m_2}{\chi^2(2m_1)/2m_1}}_{\substack{\text{independent} \\ \text{numerator and} \\ \text{denominator}}} = \frac{\lambda_x}{\lambda_y} F(2m_2, 2m_1)$$

Under H_0: $\frac{\lambda_x}{\lambda_y} = 1$ we have exactly $T \overset{H_0}{\sim} F(2m_2, 2m_1)$

But even if we didn't know this, from $(*)$ we know that under H_0, for any $\lambda_x = \lambda_y$ the distribution of T is the same. We can therefore resample - generate i.i.d. samples from any $Exp(\lambda_x)$ and $Exp(\lambda_y)$ for $\lambda_x = \lambda_y$ and simulate the distribution of T under H_0.

Bootstrap

Example: Inference about $EX = \theta$ in (possibly) Exponential distribution

$X_1, \dots, X_n$ iid from a distribution that could be exponential

$\theta = EX_1$, possibly $X_1 \sim \mathcal{E}xp(\lambda)$, $\hat{\theta} = \bar{X}$

- CLT: if $\operatorname{var} X_1$ exists, $\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{D}} N(0, \operatorname{var} X_1)$

  confidence interval: $\left[\bar{X} \mp u_{1-\alpha/2} \sqrt{\frac{\widehat{\operatorname{var}}\, X_1}{n}} \right]$

  but is only asymptotic, works only for larger n.

- $X_1 \sim \mathcal{E}xp(\lambda)$: $\theta = 1/\lambda$, $\hat{\lambda} = 1/\bar{X}$, $\sum X_i \sim \Gamma(n, 1/\lambda)$

  $\bar{X} = \frac{1}{n} \sum X_i \sim \Gamma\left(n, \frac{1}{\lambda n}\right)$ and

  $2\lambda n \bar{X} = 2\lambda \sum X_i \sim \Gamma\left(\frac{2n}{2}, 2\right) = \chi^2(2n)$ ← χ^2 with $2n$ degrees of freedom

  $\frac{2 \sum X_i}{\theta} \sim \chi^2_{2n}$ exactly, and this leads to an exact conf. interval for θ

  $$\left[\frac{2 \sum X_i}{\chi^2_{2n}(1-\alpha/2)} ; \frac{2 \sum X_i}{\chi^2_{2n}(\alpha/2)} \right]$$

  Exact, but valid only if really $X_1 \sim \mathcal{E}xp(\lambda)$

- $X_1 \sim \mathcal{E}xp(\lambda)$ + variance-stabilizing transform:

  $\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{D}} N(0, \theta^2)$

  $[g'(t)]^2 = 1/t^2 \Rightarrow g(t) = \log t$

  Δ-meta $\sqrt{n}(\log \bar{X} - \log \theta) \xrightarrow{\mathcal{D}} N(0,1)$

  $\quad \parallel$

  $N\left(0, \frac{\operatorname{var} X_1}{(EX_1)^2}\right)$

  We express the as. variance as a "sandwich estimator" $\frac{\widehat{\operatorname{var}}\, X_1}{(\widehat{EX}_1)^2}$ to make inference less sensitive to the assumption $\mathcal{E}xp(\lambda)$

Bootstrap intervals – Nonparametric Bootstrap.

Idea: If the exact $F \sim X_1, \dots, X_n$ is not known, sample from its estimator $\hat{F}_n(t) = \frac{1}{n} \sum I[X_i \le t]$ ⇒ generate i.i.d samples repeatedly (with replacement) from the set of observations $\{X_1, \dots, X_n\}$. For each $b \in \{1, \dots, B\}$ $\binom{B=}{10^4}$

- generate $X^*_{1,b}, \dots, X^*_{n,b}$ with replacement from $\{X_1, \dots, X_n\}$
- compute $\hat{\theta}^*_b$ as an estimator of θ from $X^*_{1,b}, \dots, X^*_{n,b}$
- in $\sqrt{n}(\hat{\theta} - \theta)$
  - distribution of X_1 is approximated by $\{X^*_{1,b}\}_{b=1}^B$
  - distribution of $\hat{\theta}$ is approximated by $\{\hat{\theta}^*_b\}_{b=1}^B$
  - the true value θ is replaced by $\hat{\theta}$ – an estimator of θ, and a "true value" of θ for $\hat{F}_n$ from which we sample.

overall the distribution of $H_n \sim \sqrt{n}(\hat{\theta} - \theta)$ is approximated by the sampled distribution of $H^*_n \sim \sqrt{n}(\hat{\theta}^*_b - \hat{\theta})$ and the quantiles of H_n are approximated by those of H^*_n, for B large enough.

Here we approximate the distribution of $\sqrt{n}(\bar{X} - \theta)$

- standard bootstrap : compute quantiles of $\{\sqrt{n}(\bar{X}^*_b - \bar{X})\}_{b=1}^B$ and use them as quantiles of $\sqrt{n}(\bar{X} - \theta)$. That is, for example

$$\sqrt{n}(\bar{X} - \theta) \le q_{1-\alpha/2} \approx q^*_{1-\alpha/2}$$

$$\bar{X} - \frac{q^*_{1-\alpha/2}}{\sqrt{n}} \le \theta \quad \Rightarrow \quad \left[\bar{X} - \frac{q^*_{\alpha/2}}{\sqrt{n}}, \bar{X} - \frac{q^*_{1-\alpha/2}}{\sqrt{n}}\right]$$

for q^*_α the estimated α-quantile of $\{\sqrt{n}(\bar{X}^*_b - \bar{X})\}_{b=1}^B$.

→ Standard non-parametric bootstrap

UNIVERSITÉ DE GENÈVE
GENEVA SCHOOL OF ECONOMICS AND MANAGEMENT

CiCG
CENTRE INTERNATIONAL DE CONFÉRENCES GENÈVE

- Studentized nonparametric bootstrap.

Suppose that $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{D} N(0, \sigma^2)$. If we can estimate σ^2, we have $\sqrt{n}\frac{(\hat{\theta}-\theta)}{\hat{\sigma}} \xrightarrow{D} N(0,1)$ and we can approximate $\sqrt{n}\frac{(\hat{\theta}-\theta)}{\hat{\sigma}}$ by resamples $\left\{\sqrt{n}\frac{(\hat{\theta}^*_b-\hat{\theta})}{\hat{\sigma}^*_b}\right\}_{b=1}^{B}$ for $\hat{\sigma}^*_b$ an estimate of $\hat{\sigma}$ for the b-th resample. In our case $\sigma^2 = \mathrm{var}\, X_1$ is estimated by $(\hat{\sigma}^*_b)^2 = \mathrm{var}\{X^*_{1,b}, \dots, X^*_{n,b}\}$, and the distribution of $\sqrt{n}\frac{\hat{\theta}-\theta}{\hat{\sigma}} = \sqrt{n}\frac{\bar{X}-\theta}{\sqrt{\widehat{\mathrm{var}}\, X_1}}$ is approximated by $\left\{\sqrt{n}\frac{\bar{X}^*_b-\bar{X}}{\sqrt{\mathrm{var}\{X_{i,b}\}_{i=1}^{n}}}\right\}$.

Usually it is good to studentize if we can because then the distribution of the quantity we approximate depends on θ ~~much~~ less (ideally, it does not depend on θ at all)

- Studentized NP bootstrap + variance stabilization

analogously for $\sqrt{n}(\log\bar{X} - \log\theta)$

- Diagnostics: if really $X_1 \sim Exp(\lambda)$ then $\begin{matrix} EX=\theta \\ \mathrm{var}\,X=\theta^2 \end{matrix} \Rightarrow$ $EX = \sqrt{\mathrm{var}\,X}$. This can be checked when EX and $\mathrm{var}\,X$ are approximated by EX^*_b and $\mathrm{var}\,X^*_b$, respectively. If these depart a lot, the original distribution was likely not exponential. But, it's not a test, of course.

diagnostics: if $X \sim \mathcal{E}xp(\lambda)$ $EX = 1/\lambda$ $\operatorname{var} X = 1/\lambda^2$ $\to EX = \text{s.d}(X)$

if $X \sim \mathcal{E}xp(\lambda)$ $E \log \bar{X} \approx \log 1/\lambda$ $\operatorname{var} \bar{X} \approx \frac{1}{m}\frac{\operatorname{var} X}{(EX)^2} = \frac{1}{m}$ ~ constant

---

(6)

Parametric bootstrap. We know that $F_\theta \sim X$, $\theta \in \Theta$ but don't know θ.
$\to$ estimate θ by $\hat{\theta}_m$ and generate replicates from $F_{\hat{\theta}_m}$.

Ex 44,45: $X \sim \mathcal{U}([0,\theta])$ $\theta > 0$ $\hat{\theta}_m = X_{(m)}$ $m(\theta - \hat{\theta}_m) \overset{as}{\sim} \mathcal{E}xp(1/\theta)$

But basic NP bootstrap generates $\hat{\theta}_b^* = X_{(m)}$ a lot $\Rightarrow$

$P(X^*_{(m)} = X_{(m)} \mid X) = 1 - \left(\frac{m-1}{m}\right)^m \xrightarrow[m\to\infty]{} 1 - e^{-1}$

for $R_m^* = m(\hat{\theta} - \hat{\theta}_b^*)$ $P(R_m^* = 0 \mid X) \sim 1 - e^{-1}$ for m large.

$\to$ generate X_b^* directly from $\mathcal{U}([0, \hat{\theta}_m])$ and run procedure.

Goodness-of-fit testing: $H_0: X \sim F_\theta$ for some $\theta \in \Theta$ $H_1: X \nsim F_\theta$ for any $\theta \in \Theta$

$KS_m^\theta = \sup_x |F_m(x) - F_\theta(x)|$ estimate θ by $\hat{\theta}_m$ $\to$ use $KS_m^{\hat{\theta}_m}$ dist.
to the nearest element of Θ. can't use the limiting distribution – that is based on single θ. $\to$ generate replicates from $F_{\hat{\theta}_m}$ and resample KS_m^*

Two-sample problems: $X_1 .. X_m \sim F_1$ $Y_1 .. Y_m \sim F_2$ $H_0: EX = EY$ $H_1: EX \neq EY$

- parametric bootstrap approach resample from $N(0, S_x^2)$, $N(0, S_y^2)$ and evaluate a test statistic
- non-parametric resample from $X_i - \bar{X}$, $Y_j - \bar{Y}$ and estimate from a t-statistic

Permutation test of independence: $\binom{X_1}{Y_1} - \binom{X_m}{Y_m} \sim F_{X,Y}$ H_0: X is indep of Y

under H_0 if one permutes only Y_i and keeps X_i fixed to get $\binom{X_i^*}{Y_i^*}$ then still X_i^* is indep. of Y_i^*. $\Rightarrow$ evaluate

Simulation χ^2 test of independence $X_1 \sim \text{Mult}(m_1, p_{11} .. p_{1k})$
$\vdots$
$X_3 \sim \text{Mult}(m_3; p_{31} .. p_{3k})$ $p_j = (p_{j1} .. p_{jk})$

$H_0: p_i = p_j \; \forall i \neq j$ $X_1 \sim X_{11} .. X_{m,1}$ individual observations.

under H_0 pool all observation $\sum_{i=1}^{3} m_i$ and resample (permute) this order while keeping m_i observations to X_i^* $\Rightarrow$ invariant under H_0

NMST434 Ex. session 10: Resampling II 2018/2019 (7)

Ex: Two sample problem

$X_1, \dots, X_{m_1} \sim F$ independent $\quad H_0: F \equiv G \quad$ Under H_0, $Z = (X_1, \dots, X_{m_1}, Y_1, \dots, Y_{m_2}) \overset{iid}{\sim} F$
$Y_1, \dots, Y_{m_2} \sim G \quad H_1: \neg H_0 \quad$ and Z is invariant with respect to permut.

Formally, $Z \mid Z_{(\cdot)} \sim$ Unif (all permutations of $Z_{(\cdot)}$) (ordered sample Z) "generate" data under H_0

In analogy with nonparametric Bootstrap: permute Z to get Z^*, take X^* to be the first m_1 elements of Z^*, Y^* the other m_2 elements and approximate the distribution of a test statistic T under H_0 by $T^* = T(X^*, Y^*)$

Permutation tests

Kolmogorov-Smirnov test $\quad T_m = \sup_{x \in \mathbb{R}} | \hat{F}_{m_1}(x) - \hat{G}_{m_2}(x) |$

χ^2-test of independence $\quad \binom{X_1}{Y_1}, \dots, \binom{X_m}{Y_m} \sim F$ independent, H_0: X is independent of Y
$H_1: \neg H_0$

Under H_0 the law of $Z = (X_1, \dots, X_m, Y_1, \dots, Y_m)$ is invariant w.r.t. permutations of Ys (with X's fixed) $\rightsquigarrow$ generate new data by permuting Y's

X_i = value of data, Y_i = sample indicator in k sample problem $\Rightarrow$ X is independent of Y iff all X_i have the same distribution $\rightsquigarrow$ permutation tests in k-sample problems permutations of ~~tablets~~ labels. For $X_i \sim$ Multinomial this leads to a permutation χ^2-test of independence.

Ex. Bootstrap in linear models $\quad \binom{X_1}{Y_1}, \dots, \binom{X_m}{Y_m}$ iid, $Y_i = \beta' X_i + \varepsilon$, $\varepsilon \perp X$, $E[\varepsilon] = 0$

- nonparametric bootstrap: resample directly from $\binom{X_1}{Y_1}, \dots, \binom{X_m}{Y_m}$
- model-based bootstrap: estimate $\hat\beta$, $\hat\varepsilon_i := \dfrac{Y_i - \hat\beta' X_i}{\sqrt{1 - h_{ii}}}$ standardised residuals

so that $\hat\varepsilon$ are centred with unit variance, resample from $\hat\varepsilon$ to get ε^* and generate new data as $Y_i^* = \hat\beta' X_i + \varepsilon_i^*$, $\binom{X_1}{Y_1^*}, \dots, \binom{X_m}{Y_m^*}$

- Works if the model is correct but also under fixed design.
- NP-Bootstrap works even if model is not true, but only in random design.

Ex: Durbin-Watson's test: linear model, indices have meaning, ε may be a time series.
$\varepsilon \sim AR(1)$, $\rho = \mathrm{cor}(\varepsilon_i, \varepsilon_{i+1} \mid X)$, $H_0: \rho = 0$
$H_1: \rho \neq 0$

$U = (U_1, \dots, U_m)$ residuals $\quad DW = \dfrac{\sum_{i=2}^m (U_i - U_{i-1})^2}{\sum_{i=1}^m U_i^2} \approx 2(1 - \hat\rho)$ for $\hat\rho$ estimator of ρ.

Also under H_0, law of DW depends on the design matrix X $\rightsquigarrow$ bootstrap.

Ex: Bootstrap in AR(1) $\quad X_t = a X_{t-1} + \varepsilon_t$, $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ independent of X, $a \in (-1, 1)$
data are not iid, cannot resample directly from X
model-based bootstrap $\quad \hat\varepsilon_t := X_t - \hat a X_{t-1}$, $\hat\varepsilon$ are approximately iid
$\longrightarrow$ resample from $\hat\varepsilon$ to get ε^* and set $X_t^* = \hat a X_{t-1}^* + \varepsilon_t^*$.

Ex 10) Permutation test of independence

Ex 49) k independent random samples (wlog k = 3)

$X_1 .. X_{m_1} \sim F_X$ $H_0: F_X = F_Y = F_Z$ $H_1: \neg H_0$

$Y_1 .. Y_{m_2} \sim F_Y$ under H_0, all observations are as iid, and any permutation has the same probability / original sample

For a test statistic T if H_0 is true, any permutation of the pooled sample should give a test statistic with the same distribution (T_b^*)

→ compute T

→ permute pooled data and evaluate T_b^*

→ find p-value by comparing T and $\{T_b^*\}_{b=1}^B$

Ex: Bootstrap in linear models

$\binom{X_1}{Y_1} .. \binom{X_m}{Y_m}$ iid $Y_i = \beta X_i + \varepsilon_i$ for ε_i independent of X_i and $E(\varepsilon_i | X_i) = 0$

→ nonparametric bootstrap: resample directly from $\binom{X_i}{Y_i}$

→ model-based bootstrap: estimate β by $\hat\beta$, let $\hat\varepsilon_i = \frac{Y_i - \hat\beta x_i}{\sqrt{1-h_{ii}}}$ standardized residuals ($\hat\varepsilon_i$ are centred, with variance σ^2), resample from $\hat\varepsilon_i$ to get ε_i^* and generate new response variables as $Y_i^* = \hat\beta X_i + \varepsilon_i^*$

Bootstrap sample is given by $\binom{X_1}{Y_1^*} .. \binom{X_m}{Y_m^*}$

- model-based bootstrap works if the assumed model is true also in fixed design.
- nonpar. bootstrap works always for random design, even if the model is not true.

Ex: Durbin-Watson's test. linear model, order of indices may have meaning, ε_i may be a time series. $\varepsilon_i \sim AR(1)$, $\rho = \mathrm{cor}(\varepsilon_i, \varepsilon_{i+1} | X)$

$H_0: \rho = 0$ $H_1: \rho \neq 0$ for $U = (U_1 .. U_n)^T$ residuals,

$$DW = \frac{\sum_{i=2}^n (U_i - U_{i-1})^2}{\sum_{i=1}^n U_i^2} \approx 2(1-\hat\rho)$$ for $\hat\rho$ the an estimator of ρ

Distribution of DW under H_0 depends on $\underline{X}$. → needed bootstrap.

Ex 44) Bootstrap in autoregression $x_t = a x_{t-1} + \varepsilon_t$ $\varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ indep. of x

$a \in (-1,1)$, $t \in \mathbb{Z}$ not iid $x_1, x_2 ..$ time series - can't use nonparametric resampling; model-based bootstrap $\hat\varepsilon_t = x_t - \hat a x_{t-1}$ then $\hat\varepsilon_t$ are approximating an iid sequence of ε_i → resample ε_t^* from $\hat\varepsilon_t$ and set $x_t^* = \hat a x_{t-1}^* + \varepsilon_t^*$. Then $\{x_t^*\}$ is the bootstrapped sample.

Ex 50) variance estimation in bootstrap - counterexample

bootstrapped distribution approximates the distribution of a finite sample size version of the statistic, not the asymptotic distribution