ΣΤΑΤΙΣΤΙΚΗ ΙΙΙ

Θεωρία 02

Δήμητρα Κυριακοπούλου

Τμήμα Οικονομικών Επιστημών

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

- Συνήθως το κριτήριο της αμεροληψίας καθορίζει υποκλάσεις του συνόλου των εκτιμητριών.

- Για το λόγο αυτό, θα ήταν επιθυμητό να χρησιμοποιήσουμε παράλληλα και ένα δεύτερο κριτήριο βελτιστότητας, το οποίο να περιορίζει τις υποψήφιες «βέλτιστες» στατιστικές συναρτήσεις.

- Αυτό το κριτήριο καλείται αμερόληπτη εκτιμήτρια ελάχιστης διακύμανσης, το οποίο είναι το αντικείμενο μελέτης αυτής της διάλεξης.

Αμερόληπτη Ελάχιστης Διακύμανσης (ΑΕΔ) Εκτιμήτρια

- Μία αμερόληπτη εκτιμήτρια $\hat{\theta}$ καλείται ομοιόμορφα Αμερόληπτη Ελάχιστης Διακύμανσης (ΑΕΔ) - minimum variance unbiased estimator - αν έχει τη μικρότερη διακύμανση μεταξύ όλων των αμερόληπτων εκτιμητών $\tilde{\theta}$, για κάθε θ:

$$E\left(\hat{\theta}\right) = \theta$$

και

$$Var\left(\hat{\theta}\right) \leq Var\left(\tilde{\theta}\right),$$

$$\forall \tilde{\theta}: \ E\left(\tilde{\theta}\right) = \theta.$$

- Πώς γνωρίζουμε ότι υπάρχει τέτοια αμερόληπτη εκτιμήτρια και κυρίως με ποιο τρόπο μπορούμε να την ανακαλύψουμε;

- Η απάντηση είναι με ένα θεωρητικό κατώτερο όριο (κάτω φράγμα) της διακύμανσης όλων των αμερόληπτων εκτιμήτριων.

Ανισότητα Cramer-Rao

- Με τη μέθοδο Cramer-Rao βρίσκεται ένα κάτω φράγμα (το οποίο καλούμε CR bound) της διακύμανσης της εκτιμήτριας:
  - αν $E\left(\hat{\theta}\right) = \theta$, τότε η διακύμανση της εκτιμήτριας δίνεται από τη σχέση:

$$Var\left(\hat{\theta}\right) \geq \frac{1}{nI\left(\theta\right)},$$

όπου:

$$\begin{aligned} I\left(\theta\right) &= E\left[\left(\frac{\partial \log f\left(x;\theta\right)}{\partial\theta}\right)^2\right] \\ &= -E\left[\frac{\partial^2 \log f\left(x;\theta\right)}{\partial\theta^2}\right] \end{aligned}$$

και $I\left(\theta\right)$ καλείται <u>πληροφοριακός αριθμός κατά Fisher (Fisher Information)</u> της παραμέτρου θ.

Απόδειξη:

- Έστω τυχαίο δείγμα $(X_1, X_2, ..., X_n)$ με κατανομή $f(x;\theta)$.

- Η από κοινού κατανομή του τυχαίου δείγματος συμβολίζεται με $L(x;\theta)$ και ισχύει ότι:

$$L(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- Επίσης, φέρει την εξής ιδιότητα:

$$\int_{\mathbb{R}^{\ltimes}} L(x;\theta)\, dx = 1.$$

- Επειδή $\hat{\theta}$ είναι αμερόληπτη εκτιμήτρια του θ, ισχύει ότι:

$$E\left(\hat{\theta}\right) = \theta \Rightarrow$$

$$\int_{\mathbb{R}^{\ltimes}} \theta(X_1, X_2, ..., X_n)\, L(x;\theta)\, dx = \theta$$

• Ορίζουμε μια νέα τυχαία μεταβλητή:

$$
\begin{aligned}
U \equiv \sum_{i=1}^{n} U_i \equiv \sum_{i=1}^{n} \frac{\partial \log f\left(x_i;\theta\right)}{\partial \theta} \\
= \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \log f\left(x_i;\theta\right) \\
= \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f\left(x_i;\theta\right) \\
= \frac{\frac{\partial}{\partial \theta} \prod_{i=1}^{n} f\left(x_i;\theta\right)}{\prod_{i=1}^{n} f\left(x_i;\theta\right)}.
\end{aligned}
$$

- Η μέση τιμή της δίνεται από:

$$
\begin{aligned}
E(U) = \sum_{i=1}^{n} E(U_i) &= n \int_{\mathbb{R}} \frac{\partial \log f(x;\theta)}{\partial \theta} f(x;\theta)\, dx \\
&= n \int_{\mathbb{R}} \frac{\partial f(x;\theta)}{\partial \theta} dx \\
&= n \frac{\partial}{\partial \theta} \int_{\mathbb{R}} f(x;\theta)\, dx \\
&= n \frac{\partial}{\partial \theta} (1) = 0.
\end{aligned}
$$

- Και η διακύμανσή της:

$$Var\left(U\right)=\sum_{i=1}^{n}Var\left(U_i\right)=nE\left[\left(\frac{\partial\log f\left(x_i;\theta\right)}{\partial\theta}\right)^2\right].$$

- Λόγω της ιδιότητας της αμεροληψίας, δείξαμε πιο πριν ότι:

$$\int_{\mathbb{R}^{\ltimes}}\theta\left(X_1,X_2,...,X_n\right)L\left(x;\theta\right)dx=\theta$$

και παραγωγίζοντας ως προς θ έχουμε:

$$E\left(\hat{\theta}\cdot U\right)=1\Rightarrow$$
$$Cov\left(\hat{\theta},U\right)=1.$$

- Όμως ξέρουμε ότι ισχύει η γνωστή από τις πιθανότητες <u>ανισότητα συνδιακύμανσης:</u>

$$Cov\left(\hat{\theta},U\right)^2\leq Var\left(\hat{\theta}\right)\cdot Var\left(U\right)\Rightarrow$$
$$Var\left(\hat{\theta}\right)\geq\frac{1}{nE\left[\left(\frac{\partial\log f(x_i;\theta)}{\partial\theta}\right)^2\right]}.$$

• Τέλος, είναι σημαντικό να αποδείξουμε τη σχέση:

$$E\left[\left(\frac{\partial \log f(x;\theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}\right].$$

• Έχουμε ότι:

$$\frac{\partial \log f(x_i;\theta)}{\partial \theta} = \frac{f'(x_i;\theta)}{f(x_i;\theta)}$$

και παραγωγίζοντας ξανά ως προς θ, βρίσκουμε:

$$\begin{aligned}\frac{\partial^2 \log f(x_i;\theta)}{\partial \theta^2} &= \frac{f''(x_i;\theta)\, f(x_i;\theta) - f'(x_i;\theta)\, f'(x_i;\theta)}{[f(x_i;\theta)]^2} \\ &= \frac{f''(x_i;\theta)}{f(x_i;\theta)} - \left[\frac{\partial \log f(x_i;\theta)}{\partial \theta}\right]^2\end{aligned}$$

- Αν πάρουμε τις μέσες τιμές των δύο μελών, έχουμε:

$$E\left[\frac{\partial^2 \log f\left(x_i;\theta\right)}{\partial\theta^2}\right] = E\left[\frac{f''\left(x_i;\theta\right)}{f\left(x_i;\theta\right)}\right] - E\left[\frac{\partial \log f\left(x_i;\theta\right)}{\partial\theta}\right]^2$$

και ισχύει:

$$E\left[\frac{f''\left(x_i;\theta\right)}{f\left(x_i;\theta\right)}\right] = 0,$$

επειδή

$$\frac{\partial^2}{\partial\theta^2}\left(\int f\left(x_i;\theta\right)dx\right) = 0.$$

- Επομένως:

$$E\left[\left(\frac{\partial \log f\left(x;\theta\right)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \log f\left(x;\theta\right)}{\partial\theta^2}\right].$$

- <u>Παράδειγμα 1:</u>

  Έστω τυχαίο δείγμα $(X_1, X_2, ..., X_n)$ από πληθυσμό με συνάρτηση κατανομής την Κανονική Κατανομή $N\left(\mu, \sigma^2\right)$:

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Τότε ισχύει:

$$\log f = \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{(x-\mu)^2}{2\sigma^2}$$

και θεωρώντας άγνωστη μόνο την παράμετρο μ:

$$\begin{aligned} \frac{\partial}{\partial\mu} \log f &= -2\frac{(x-\mu)}{2\sigma^2}(-1) \\ &= \frac{x-\mu}{\sigma^2}, \\ \frac{\partial^2}{\partial\mu^2} \log f &= -\frac{1}{\sigma^2}. \end{aligned}$$

Επίσης:

$$E\left(\frac{\partial^2}{\partial\mu^2}\log f\right) = -\frac{1}{\sigma^2} \Rightarrow$$
$$I(\mu) = \frac{1}{\sigma^2}$$

Και από το οποίο προκύπτει ότι:

$$CR = \frac{1}{n\frac{1}{\sigma^2}} = \frac{\sigma^2}{n}.$$

Οπότε για $\theta = \{\mu\}$, ισχύει:

$$Var\left(\hat{\theta}\right) \geq \frac{1}{nI(\theta)} = \frac{\sigma^2}{n} = Var\left(\bar{X}\right).$$

- Μια εκτιμήτρια της οποίας η διακύμανση είναι τόσο μικρή όσο το κάτω φράγμα CR καλείται **αποτελεσματική** (efficient).
- Στο παραπάνω παράδειγμα ο δειγματικός μέσος είναι αποτελεσματικός.
- Για το λόγο αυτό, καλείται η καλύτερη αμερόληπτη εκτιμήτρια της παραμέτρου μ.

- Παράδειγμα 2:

  Έστω τυχαίο δείγμα $(X_1, X_2, ..., X_n)$ που ακολουθεί την κατανομή βήτα $\beta(\theta, 1)$. Να βρεθεί αμερόληπτη εκτιμήτρια ομοιόμορφα ελάχιστης διακύμανσης (ΑΕΔ) της συνάρτησης $g(\theta) = \frac{1}{\theta}$.

  Η συνάρτηση πυκνότητας πιθανότητας της κατανομής $\beta(\theta, 1)$ δίνεται από τη σχέση:

$$f(x; \theta) = \theta x^{\theta - 1}, \quad 0 < x < 1, \quad \theta > 0.$$

  Η από κοινού κατανομή του τυχαίου δείγματος $L(x; \theta)$ είναι:

$$L(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \theta x^{\theta - 1} = \theta^n \prod_{i=1}^{n} x_i^{\theta - 1}.$$

Λογαριθμίζοντας την προηγούμενη σχέση προκύπτει:

$$\log L(x;\theta) = n\log\theta + (\theta - 1)\sum_{i=1}^{n}\log x_i.$$

Και παραγωγίζοντας ως προς θ έχουμε ότι:

$$\begin{aligned}\frac{\partial \log L(x;\theta)}{\partial\theta} &= \frac{n}{\theta} + \sum_{i=1}^{n}\log x_i \\ &= -n\left(-\frac{\sum_{i=1}^{n}\log x_i}{n} - \frac{1}{\theta}\right).\end{aligned}$$

Για τη συνέχεια, θα χρειαστούμε το παρακάτω Πόρισμα:

Αν ισχύει ότι:

$$\frac{\partial \log L\left(x;\theta\right)}{\partial\theta} = k\left(\theta\right)\left(U\left(x\right) - g\left(\theta\right)\right),$$

τότε η στατιστική $U\left(x\right)$ είναι εκτιμήτρια ΑΕΔ για τη συνάρτηση $g\left(\theta\right)$ με

$$Var\left[U\left(x\right)\right] = \frac{g'\left(\theta\right)}{k\left(\theta\right)}.$$

Οπότε, η στατιστική:

$$-\frac{\sum_{i=1}^{n} \log x_i}{n}$$

είναι εκτιμήτρια ομοιόμορφα αμερόληπτη ελάχιστης διακύμανσης (ΑΕΔ) της συνάρτησης

$$g(\theta) = \frac{1}{\theta}.$$

Η διακύμανση της ισούται με:

$$\begin{aligned} Var\left[U(x)\right] &= \frac{g'(\theta)}{k(\theta)} \\ &= \frac{\left(\frac{1}{\theta}\right)'}{-n} \\ &= \frac{1}{n\theta^2}. \end{aligned}$$

- <u>Παράδειγμα 3:</u>

Έστω τυχαίο δείγμα $(X_1, X_2, ..., X_n)$ i.i.d. που ακολουθεί την κατανομή Poisson με παράμετρο λ.

Να δείξετε ότι ο δειγματικός μέσος $\bar{X}$ είναι αμερόληπτη εκτιμήτρια ομοιόμορφα ελάχιστης διακύμανσης για την παράμετρο λ.

Γνωρίζουμε ότι αν $X \sim Poisson\,(\lambda)$,

$$f\,(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!},$$

με $E\,(X) = \lambda$ και $Var\,(X) = \lambda$.

Οπότε:

$$E\left(\bar{X}\right) = \lambda$$
$$Var\left(\bar{X}\right) = \frac{\lambda}{n},$$

και άρα ο δειγματικός μέσος $\bar{X}$ είναι αμερόληπτη εκτιμήτρια του λ.

Για να δείξουμε ότι είναι και ομοιόμορφα ελάχιστης διακύμανσης, πρέπει να ισχύει ότι:

$$Var\left(\bar{X}\right) = \frac{1}{I\left(\lambda\right)},$$

όπου $I\left(\lambda\right)$ ο Fisher information για την παράμετρο λ.

Το πρώτο βήμα είναι να υπολογίσουμε τον λογάριθμο της από κοινού κατανομής:

$$\begin{aligned}
\ell \log \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} &= \log e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \frac{1}{x_i!} \\
&= -n\lambda + \sum_{i=1}^{n} x_i \log\lambda - log \prod_{i=1}^{n} x_i! \\
&= -n\lambda + n\bar{x}\log\lambda - log \prod_{i=1}^{n} x_i!
\end{aligned}$$

Στη συνέχεια, θα υπολογίσουμε τις παραγώγους:

$$\frac{\partial \ell}{\partial \lambda} = -n + \frac{n\bar{x}}{\lambda} \Rightarrow$$
$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{n\bar{x}}{\lambda^2}.$$

Οπότε:

$$\begin{aligned} I(\lambda) &= -E\left(\frac{\partial^2 \ell}{\partial \lambda^2}\right) \\ &= E\left(\frac{n\bar{X}}{\lambda^2}\right) \\ &= \frac{n}{\lambda^2}E\left(\bar{X}\right) \\ &= \frac{n}{\lambda}. \end{aligned}$$

Και ισχύει ότι:

$$Var\left(\bar{X}\right) = \frac{\lambda}{n} = \frac{1}{I(\lambda)}$$

$\Rightarrow$ $\bar{X}$ είναι αμερόληπτη εκτιμήτρια ομοιόμορφα ελάχιστης διακύμανσης του λ.

- Σε αυτό το σημείο αξίζει να σημειωθεί ότι μπορεί σε κάποιες περιπτώσεις να μην θέλουμε να περιοριστούμε σε αμερόληπτες εκτιμήτριες σαν κριτήριο βελτιστότητας των διαφόρων εκτιμητριών.

- Η σύγκριση δύο **αμερόληπτων** εκτιμητριών γίνεται με βάση τη διακύμανση και <u>προτιμάται η εκτιμήτρια με τη μικρότερη διακύμανση.</u>

- Η σύγκριση μεταξύ δύο **μεροληπτικών** εκτιμητριών γίνεται με βάση το <u>μέσο τετραγωνικό σφάλμα (mean squared error - MSE):</u>

$$
\begin{aligned}
E\left[\left(\hat{\theta}-\theta\right)^2\right] &:= MSE\left(\hat{\theta}\right) \\
&= Bias^2\left(\hat{\theta}\right) + Var\left(\hat{\theta}\right) \\
&= \left(E\left(\hat{\theta}\right)-\theta\right)^2 + Var\left(\hat{\theta}\right).
\end{aligned}
$$

- Παράδειγμα:

  Έστω οι αμερόληπτες και ασυσχέτιστες εκτιμήτριες $\hat{\theta}_1$ και $\hat{\theta}_2$ της παραμέτρου θ. Να βρεθεί γραμμικός συνδυασμός:

$$\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$$

  τέτοιος ώστε η εκτιμήτρια $\hat{\theta}$ να είναι αμερόληπτη και να έχει την ελάχιστη διακύμανση. Έχουμε ότι:

$$\begin{aligned} E\left(\hat{\theta}\right) &= c_1 E\left(\hat{\theta}_1\right) + c_2 E\left(\hat{\theta}_2\right) \\ &= c_1\theta + c_2\theta \\ &= \left(c_1 + c_2\right)\theta. \end{aligned}$$

Για να είναι αμερόληπτη εκτιμήτρια θα πρέπει:

$$
\begin{aligned}
E\left(\hat{\theta}\right) &= \theta \Rightarrow \\
\left(c_1 + c_2\right)\theta &= \theta \Rightarrow \\
c_1 + c_2 &= 1 \Rightarrow \\
c_2 &= 1 - c_1.
\end{aligned}
$$

Άρα:

$$
\begin{aligned}
\hat{\theta} &= c_1\hat{\theta}_1 + \left(1 - c_1\right)\hat{\theta}_2 \Rightarrow \\
Var\left(\hat{\theta}\right) &= c_1^2 Var\left(\hat{\theta}_1\right) + \left(1 - c_1\right)^2 Var\left(\hat{\theta}_2\right) \\
&\quad + 2c_1\left(1 - c_1\right) Cov\left(\hat{\theta}_1, \hat{\theta}_2\right) \\
&= c_1^2\sigma_1^2 + \left(1 - c_1\right)^2\sigma_2^2 + 2c_1\left(1 - c_1\right)\sigma_{12}.
\end{aligned}
$$

Έστω $\sigma_{12} = 0$. Τότε:

$$\frac{\partial Var\left(\hat{\theta}\right)}{\partial c_1} = 2c_1\sigma_1^2 - 2\left(1 - c_1\right)\sigma_2^2$$

$$\frac{\partial Var\left(\hat{\theta}\right)}{\partial c_1} = 0 \iff c_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Άρα:

$$\hat{\theta} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\hat{\theta}_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\hat{\theta}_2.$$

Η διακύμανση της αμερόληπτης εκτιμήτριας είναι:

$$
\begin{aligned}
Var\left(\hat{\theta}\right) &= \left(\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^2 \sigma_1^2 + \left(\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2 \sigma_2^2 \\
&= \frac{\sigma_2^4\sigma_1^2 + \sigma_1^4\sigma_2^2}{\left(\sigma_1^2+\sigma_2^2\right)^2} \\
&= \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \\
&= \frac{Var\left(\hat{\theta}_1\right)Var\left(\hat{\theta}_2\right)}{Var\left(\hat{\theta}_1\right)+Var\left(\hat{\theta}_2\right)}.
\end{aligned}
$$

Μέσο Τετραγωνικό Σφάλμα (MSE)

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= Var\left(\hat{\theta}\right) + \left(E\left(\hat{\theta}\right) - \theta\right)^2 \\
&= Var\left(\hat{\theta}\right) + \left(Bias\left(\hat{\theta}\right)\right)^2
\end{aligned}
$$

- Σχετική αποτελεσματικότητα ανάμεσα σε δύο εκτιμήτριες $\hat{\theta_1}$ και $\hat{\theta_2}$:

$$
MSE\left(\hat{\theta_1}\right) \leq MSE\left(\hat{\theta_2}\right)
$$

- <u>Απόδειξη:</u>

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)+E\left(\hat{\theta}\right)-\theta\right)^2\right] \\
& E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^2 \\
&+2\left(E\left(\hat{\theta}\right)-\theta\right)E\left[\hat{\theta}-E\left(\hat{\theta}\right)\right].
\end{aligned}
$$

Όμως ισχύει ότι:

$$\hat{\theta}-E\left(\hat{\theta}\right)=0,$$

οπότε:

$$
\begin{aligned}
MSE\left(\hat{\theta}\right) &= E\left[\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right]+\left(E\left(\hat{\theta}\right)-\theta\right)^2 \\
&= Var\left(\hat{\theta}\right)+\left(Bias\left(\hat{\theta}\right)\right)^2.
\end{aligned}
$$

- Παράδειγμα:

Έστω X_1, X_2, X_3 τυχαίο δείγμα από πληθυσμό της τυχαίας μεταβλητής X με κατανομή:

$$p = P(X = 1)$$
$$1 - p = P(X = 0).$$

Έστω οι εκτιμήτριες:

$$\hat{p_1} = \frac{X_1 + X_2 + X_3}{3},$$
$$\hat{p_2} = \frac{1}{2}.$$

Ισχύει ότι:

$$E(\hat{p_1}) = p,$$
$$E(\hat{p_2}) = \frac{1}{2}.$$

Άρα η εκτιμήτρια $\hat{p_1}$ είναι αμερόληπτη για κάθε p, ενώ η εκτιμήτρια $\hat{p_2}$ είναι αμερόληπτη για $p = \frac{1}{2}$ μόνο.

Συνέχεια του Παραδείγματος:

$$Var\left(\hat{p_1}\right) = \frac{p\left(1-p\right)}{3},$$
$$Var\left(\hat{p_2}\right) = 0.$$

Συνεπώς:

$$MSE\left(\hat{p_1}\right) = \frac{p\left(1-p\right)}{3},$$
$$MSE\left(\hat{p_2}\right) = \left(p - \frac{1}{2}\right)^2.$$

- Αν $p = \frac{3}{4}$ ή $p = \frac{1}{4}$:

$$MSE\left(\hat{p_1}\right) = MSE\left(\hat{p_2}\right) = \frac{1}{16}.$$

- Αν $\frac{1}{4} < p < \frac{3}{4}$:

$$MSE\left(\hat{p_2}\right) < MSE\left(\hat{p_1}\right).$$

- Αν $p < \frac{1}{4}$ ή $p > \frac{3}{4}$:

$$MSE\left(\hat{p_2}\right) > MSE\left(\hat{p_1}\right).$$