Poprzedni wykład

- Wzór na przenoszenie błędów $(X_1,X_2,..) \to (Y_1,Y_2,\ldots)$

$$cov(Y_k,Y_q)=\sum_{l,m}\left(\frac{\partial Y_k}{\partial X_l}\right)\Bigg|_{\vec{X}=E(\vec{X})}\left(\frac{\partial Y_q}{\partial X_m}\right)\Bigg|_{\vec{X}=E(\vec{X})} cov(X_l,X_m)$$

$$C(Y)=TC(X)T^T$$

- Transformacja łącznej funkcji gęstości

$$f_{V,W}(v,w)=f_{X,Y}(x(v,w),y(v,w))\left|J_{(X,Y)}\right|^{-1}$$

Przykład: transformacja Box-Mullera dająca N(0,1)

1

Dystrybuanta warunkowa

- Syt. Szukamy **dystrybuanty warunkowej zmiennej X** czyli dystrybuanty zmiennej X pod warunkiem, że zmienna losowa Y=y. Ale dla zmiennej ciągłej P(Y=y)=0 dlatego rozpatrujemy y<Y<y+dy i przejdziemy do granicy dy→0.

$$\int_{-\infty}^{x}\int_{y}^{y+dy} f_{X,Y}(u,v)\,du\,dv = F_{X|Y}(x\,|\,Y\in(y,y+dy))\cdot\int_{y}^{y+dy} f_Y(v)\,dv$$

$$\int_{-\infty}^{x} f_{X,Y}(u,y)\,du\,dy = F_{X|Y}(x\,|\,Y\in(y,y+dy))\cdot f_Y(y)\,dy \qquad /:dy \quad /\,dy\to 0$$

$$\int_{-\infty}^{x} f_{X,Y}(u,y)\,du = F_{X|Y}(x\,|\,Y=y))\cdot f_Y(y)$$

$$\boxed{F_{X|Y}(x\,|\,y)=\frac{\int_{-\infty}^{x} f_{X,Y}(u,y)\,du}{f_Y(y)}}$$

2

Warunkowa funkcja gęstości prawdopodobieństwa

$$\boxed{f_{X|Y}(x\,|\,y)=\frac{d}{dx}F_{X|Y}(x\,|\,y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}}$$

- Definicja ta prowadzi do równoważnej definicji niezależności dwóch zmiennych losowych poprzez warunek:

$$f_{X|Y}(x\,|\,y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_{X|Y}(x\,|\,y)\cdot f_Y(y)=f_{X,Y}(x,y)$$

$$f_{X|Y}(x\,|\,y)\cdot f_Y(y)=f_X(x)\cdot f_Y(y)$$

$$\boxed{\begin{array}{l} f_{X|Y}(x\,|\,y)=f_X(x) \\ F_{X|Y}(v\,|\,y)=F_X(v)\end{array}} \qquad /\int_{-\infty}^{v} dx$$

3

Warunkowa wartość oczekiwana

$$\boxed{E(X\,|\,Y=y)=\int_{-\infty}^{+\infty} x f_{X|Y}(x\,|\,y)dx}$$

- Jest to funkcja zmiennej losowej Y, zatem E(X|Y) też jest zmienną losową

$$E[E(X\,|\,Y)]=\int_{-\infty}^{+\infty} E(X\,|\,Y=y)\,f_Y(y)\,dy=$$

$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} x\,f_{X|Y}(x\,|\,y)\,dx\right]f_Y(y)\,dy=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} x\,\frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx\right]f_Y(y)\,dy=$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\,f_{X,Y}(x,y)\,dx\,dy=\int_{-\infty}^{+\infty} x\,f_X(x)\,dx=E(X)$$

ogólnie dla dowolnej funkcji h(X) $\boxed{E(h(X))=E[E(h(X)\,|\,Y)]}$

4

5

Statystyka – podstawowe pojęcia

- **Populacja generalna** – zbiór wszystkich możliwych wyników (skończony lub nieskończony).
- **Próba** – skończony zbiór doświadczeń.
- **Estymacja** – zajmuje się wnioskowaniem o własnościach populacji na podstawie próby.
  Najczęściej interesują nas wartość oczekiwana i odchylenie standardowe (lub wariancja).
- Próba prosta to ciąg niezależnych doświadczeń odnoszących się do tej samej populacji generalnej. Wynikiem n-elementowej próby jest n-wymiarowa zmienna losowa.

- **Statystyka** – funkcja zmiennych losowych obserwowanych w próbie, sama też jest zmienną losową.

6

Estymatory - podstawowe definicje

- Syt. $X_1,\ldots,X_n$ – wyniki pomiarów w próbie, mają one pewien rozkład prawdopodobieństwa (funkcję gęstości prawdopodobieństwa). Zakładamy, że ten rozkład jest taki sam dla każdego X_i, czyli wylosowanie pojedynczego X_i nie zmienia rozkładu (np. przedziałów rozkładu jednorodnego) i próba jest odpowiednio liczna. Rozkład ten zależy od jakiegoś parametru Θ (lub kilku parametrów).
- **Estymatorem** parametru Θ nazywamy statystykę o rozkładzie prawdopodobieństwa zależnym od Θ i oznaczamy $T_n(\Theta)$ lub $T_n(x_1,x_2,\ldots,x_n;\Theta)$.
  Zatem dla danego parametru może istnieć wiele estymatorów o różnych własnościach. Pożądanymi własnościami są:
- **1. Zgodność** – estymator $T_n(\Theta)$ jest zgodny gdy
  $$\forall \varepsilon > 0: \lim_{n\to\infty} P(|T_n(\Theta) - \Theta| < \varepsilon) = 1$$
  wartość estymatora (statystyki)

7

Estymatory - podstawowe definicje

- **2. Obciążenie (jego brak)** – estymator $T_n(\Theta)$ jest zmienną losową (różne próby dają różne wartości estymatora), zatem ma wartość oczekiwaną i wariancję. Obciążenie estymatora to
  $$B_n \equiv E(T_n(\Theta)) - \Theta$$
- **Estymator jest nieobciążony**, gdy niezależnie od wielkości próby
  $$\boxed{E(T_n(\Theta)) = \Theta \quad \text{czyli} \quad B_n = 0}$$
- Estymator jest obciążony gdy $B_n \neq 0$.
- **Estymator jest asymptotycznie nieobciążony** gdy
  $$\lim_{n\to\infty} B_n = 0$$
- **3. Efektywność** – estymator jest **najbardziej efektywny** gdy ma najmniejszą wariancję.
- Optymalnie estymator powinien być: zgodny, nieobciążony i najbardziej efektywny.

8

Estymatory – własności ogólne

- **Estymacja punktowa** – oszacowanie wartości parametru Θ przez podanie wartości estymatora $T_n(\Theta)$ tego parametru.
- **Estymacja przedziałowa** – podanie przedziału liczbowego, wewnątrz którego, z założonym prawdopodobieństwem, leży prawdziwa wartość parametru Θ.

Estymacja punktowa:

- Estymatorem wartości oczekiwanej jest średnia z próby:
  $$\boxed{T_n(E(X)) \equiv \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

9

Estymatory odchylenia standardowego

- Estymator zgodny, asymptotycznie nieobciążony
  $$\boxed{T_n(\sigma(X)) \equiv S(X) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}}$$
  Uwaga $S^2(X)$ jest estymatorem nieobciążonym wariancji
- Estymator zgodny, asymptotycznie nieobciążony
  $$T_n(\sigma(X)) \equiv \hat{S}(X) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2}$$
- Estymator zgodny, nieobciążony, najbardziej efektywny
  $$T_n(\sigma(X)) \equiv \tilde{S}(X) = \sqrt{\sqrt{\frac{n-1}{2}}\frac{\Gamma(0.5(n-1))}{\Gamma(0.5n)} S^2(X)}$$

10

Estymatory współczynnika korelacji

$$T_n(\rho_{X,Y}) = \frac{\sum_{i=1}^{n}(x_i - \overline{X})\cdot(y_i - \overline{Y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i - \overline{X})^2\right)\left(\sum_{i=1}^{n}(y_i - \overline{Y})^2\right)}}$$

Ten estymator jest zgodny i obciążony.

11

Estymacja przedziałowa

- **Estymacja przedziałowa** – podanie przedziału liczbowego,
  $$\left[T_n^L(\theta), T_n^P(\theta)\right]$$
  wewnątrz którego, z założonym prawdopodobieństwem $\gamma = 1-\alpha$, leży prawdziwa wartość parametru Θ.
- Przedział ten nazywamy **przedziałem ufności** dla parametru Θ na poziomie ufności $1-\alpha = \gamma$.
- Własności przedziału ufności:

  1). $P\left(T_n^L(\theta) \le \theta \le T_n^P(\theta)\right) = 1-\alpha \equiv \gamma$

  2). Końce przedziału zależą od próby i od γ, a nie zależą funkcyjnie od Θ.

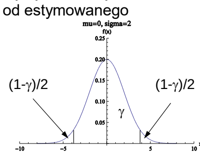
  3). Zwykle używa się $1-\alpha \approx 0.9$ (zwiększając $1-\alpha$ tracimy na dokładności, bo zwiększa się długość przedziału ufności).

12

Estymacja przedziałowa wartości oczekiwanej

- Aby znaleźć przedział ufności szukamy statystyki o znanym rozkładzie prawdopodobieństwa zależnej od estymowanego parametru Θ i próby.
- **Przypadek 1** – znamy σ(X)
- Twierdzenie: Statystyka

$$Z \equiv \frac{\overline{X} - E(X)}{\sigma(\overline{X})} = \frac{(\overline{X} - E(X))\sqrt{n}}{\sigma(X)}$$



ma rozkład normalny N(0,1). (Średnia ma rozkład N(E(X),σ²(X)/n), standaryzacja przekształca rozkład normalny w normalny).

Chcemy znaleźć przedział ufności taki, że

$$P(Z^L \le Z \le Z^P) = \gamma$$

Kwantyle rozkładu N(0,1)

Jest to spełnione przez $Z^L = Z_{\frac{1-\gamma}{2}}$ $\quad i \quad$ $Z^P = Z_{\frac{1-\gamma}{2}+\gamma=\frac{1+\gamma}{2}}$

Estymacja przedziałowa wartości oczekiwanej

- Dla rozkładu N(0,1) (z parzystości) $Z_{\frac{1-\gamma}{2}} = -Z_{\frac{1+\gamma}{2}}$, a stąd

$$P\left(-Z_{\frac{1+\gamma}{2}} \le Z \le Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(-Z_{\frac{1+\gamma}{2}} \le \frac{(\overline{X} - E(X))\sqrt{n}}{\sigma(X)} \le Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(-\frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le (\overline{X} - E(X)) \le \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(-\overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le -E(X) \le -\overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(\overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le E(X) \le \overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

- Stąd

$$\boxed{T_n^L(E(X)) = \overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \qquad T_n^P(E(X)) = \overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}}$$

Estymacja przedziałowa wartości oczekiwanej

- **Przypadek 2** – nie znamy σ(X)
- Twierdzenie: Statystyka

$$t \equiv \frac{\overline{X} - E(X)}{S(\overline{X})} = \frac{(\overline{X} - E(X))\sqrt{n}}{S(X)}$$

$$S(\overline{X}) = \sqrt{\frac{1}{n}}S(X) = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

ma rozkład t-Studenta o n-1 stopniach swobody. Tak jak w poprzednim przypadku jest to rozkład symetryczny względem x=0 i analogicznie:

$$P\left(\overline{X} - \frac{1}{\sqrt{n}}S(X)t_{\frac{1+\gamma}{2}} \le E(X) \le \overline{X} + \frac{1}{\sqrt{n}}S(X)t_{\frac{1+\gamma}{2}}\right) = \gamma$$

$t_{\frac{1-\gamma}{2}}$ $\; i \;$ $t_{\frac{1+\gamma}{2}}$ to kwantyle rozkładu t-Studenta o n-1 stopniach swobody