15 Statystyka matematyczna

Rachunek prawdopodobieństwa opracował narzędzia potrzebne do obliczania szansy uzyskania konkretnych wyników eksperymentu losowego na podstawie ogólnych praw nim rządzących. Statystyka matematyczna zajmuje się problemem odwrotnym i umożliwia wnioskowanie na temat praw ogólnych na podstawie wyników eksperymentu losowego.

Definicja 15.1 (Model statystyczny eksperymentu losowego)**.**
Model statystyczny eksperymentu losowego zakłada, że jego wyniki, $x_1, x_2, \ldots, x_n$, są realizacją zbioru zmiennych losowych $X_1, X_2, \ldots, X_n$ zwanych próbą losową.

- Model może zawierać częściową specyfikację rozkładu pr. jakiemu podlegają X_i
- Np. model zakłada, że X_i podlegają rozkładowi Poissona o nieznanych wartościach parametrów μ

W ogólnym przypadku zmienne losowe X_i mogą podlegać różnym rozkładom prawdopodobieństwa i być statystycznie zależne.

Definicja 15.2 (Próba prosta)**.**
Próbę losową nazywamy prostą, jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ podlegają temu samemu rozkładowi i są statystycznie niezależne.

Często wyniki eksperymentu losowego charakteryzujemy za pomocą jednej wartości która jest funkcją tych wyników:

$$y = h(x_1, x_2, \ldots, x_n)$$

Definicja 15.3 (Statystyka)**.**
Statystyką Y nazywamy zmienną losową realizacją której jest wartość $y = h(x_1, x_2, \ldots, x_n)$

$$Y = h(X_1, X_2, \ldots, X_n)$$

Statystyka jest zatem zmienną losową będącą funkcją próby losowej.

Przykład 15.1 (Wartość średnia)**.**
Częstą praktyką jest powtarzanie pomiaru wielokrotnie w tych samych warunkach i prezentowanie wyniku jako wartość średnią z uzyskanych pomiarów.

$$\bar{x} = h(x_1, x_2, \ldots, x_n) = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_i^n x_i}{n}$$

Statystyką której realizacją jest wartość średnia jest wartość średnia z próby:

$$\bar{X} = h(X_1, X_2, \ldots, X_n) = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_i^n X_i}{n}$$

Wyniki eksperymentu mogą posłużyć do oszacowania wielkości Θ, którą jesteśmy zainteresowani. Z reguły wielkość ta jest parametrem lub funkcją parametrów rozkładu modelowego.

Definicja 15.4 (Estymata)**.**
Estymatą wielkości Θ nazywamy wartość θ, będącą funkcją tylko wyników eksperymentu

$$\theta = h(x_1, x_2, \ldots, x_n)$$

Definicja 15.5 (Estymator)**.**
Niech $\theta = h(x_1, x_2, \ldots, x_n)$ będzie estymatą wielkości Θ. Estymata θ jest realizacją statystyki

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

zwanej estymatorem wielkości Θ.

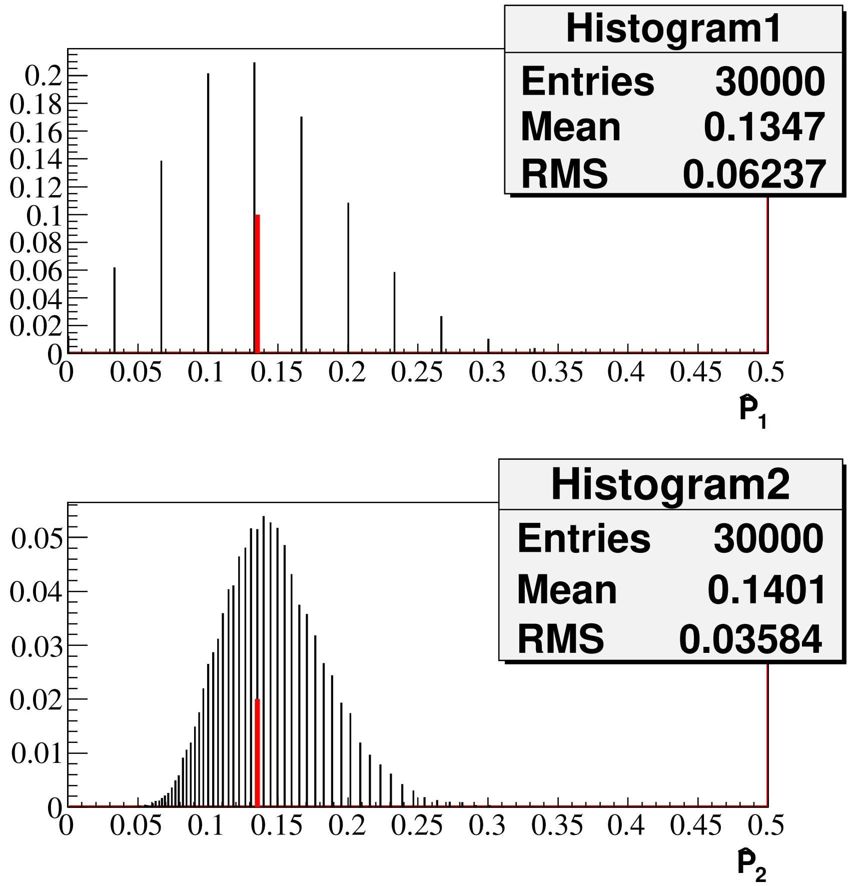
Przykład 15.2 (Estymator)**.**
Niech $x_i \quad i = 1, 2, \ldots, n$ oznaczają ilości sprzedanych samochodów w salonie z ostatnich n dni roboczych. Jeśli sprzedamy jakiś samochód to na następny dzień będziemy musieli sprowadzić brakujące modele z fabryki. Chcemy oszacować prawdopodobieństwo P, że w danym dniu nie sprzedamy żadnego samochodu.

Jak skonstruować estymator wielkości P?
Pr. P możemy oszacować na dwa sposoby:

- Jeśli przez sukces będziemy rozumieć zdarzenie polegające na braku sprzedaży w danym dniu to estymatorem, $\hat{P}_1$ może być stosunek ilości dni w których nie sprzedano żadnego samochodu do ilości wszystkich dni obserwacji:
$$\hat{P}_1 = \frac{k_0}{n} \qquad k_0 = \sum_{i:x_i=0} 1$$
- Z drugiej strony możemy założyć, że ilość sprzedanych samochodów każdego dnia podlega rozkładowi Poissona o parametrze μ, który możemy oszacować jako średnią ilość sprzedanych samochodów na jeden dzień, $\hat{\mu} = \bar{X}$, a szukane prawdopodobieństwo $P = P(k = 0) = e^{-\mu}$. Zatem
$$\hat{P}_2 = e^{-\bar{X}}$$

Rysunek 15 przedstawia wyniki symulacji Monte Carlo dla $n = 30$ i wartości x_i pochodzących z rozkładu Poissona o parametrze $\mu = 2$.
Prawdziwa wartość estymowanego parametru wynosi $P = P(k = 0) = e^{-2} \approx 0,1353$. Na podstawie symulacji Monte Carlo możemy oszacować wartość oczekiwaną oraz odchylenie standardowe estymatorów wielkości P. W przypadku estymatora $\hat{P}_1$ wartości te wynoszą $0,1354$ oraz $0,0630$ a dla estymatora $\hat{P}_2$ odpowiednio $0,1402$ oraz $0,0361$. Wartość oczekiwana estymatora $\hat{P}_1$ jest bliższa wartości prawdziwej niż estymatora $\hat{P}_2$. Z drugiej jednak strony rozrzut estymatora $\hat{P}_1$ wokół wartości oczekiwanej jest większy niż estymatora $\hat{P}_2$. Oba te parametry są podstawą oceny estymatorów.
Estymator będąc zmienną losową ma określony rozkład prawdopodobieństwa
$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$
zależny od rozkładów zmiennych losowych X_i

Przykład 15.3 (Rozkład estymatora)**.**
Jakie są rozkłady prawdopodobieństwa estymatorów $\hat{P}_1, \hat{P}_2$ z poprzedniego przykładu?
$$\hat{P}_1 = \frac{k_0}{n}$$
zmienna k_0 podlega rozkładowi dwumianowemu
$$Bin(k_0; n, P) = \begin{pmatrix} n \\ k_0 \end{pmatrix} P^{k_0}(1-P)^{n-k_0} \quad k_0 = 0, 1, \ldots n$$
Jest to rozkład dyskretny zatem rozkład estymatora też jest rozkładem dyskretnym i ma postać:
$$P(\hat{P}_1 = p_1 = \frac{k_0}{n}) = Bin(np_1; n, P) \quad p_1 = 0, 1/n, 2/n, \ldots 1$$
Powyższy rozkład dla $n = 30$ i $P = e^{-2}$ odpowiada górnemu wykresowi na rysunku 14.
$$\hat{P}_2 = e^{-\bar{X}}$$



Rysunek 14: Częstość wartości estymatora $\hat{P}_1$ (wykres górny) oraz $\hat{P}_2$ (rysunek dolny) dla próby losowej o liczebności $n = 30$ podlegających rozkładowi Poissona o parametrze $\mu = 2$. Próbę losową wygenerowano 30000 razy. Obserwowane częstości są dobrym przybliżeniem rozkładów prawdopodobieństwa jakim podlegają estymatory $\hat{P}_1$ oraz $\hat{P}_2$. Czerwoną kreską oznaczono położenie estymowanej wielkości (wartość prawdziwa)

zmienne X_i podlegają rozkładowi Poissona o parametrze

$$\mu = -\ln P; \qquad (P = e^{-\mu})$$

$$P(X_i = k) = e^{-\mu}\frac{\mu^k}{k!} \qquad k = 0, 1, \ldots, \infty$$

$Z = \sum_i^n X_i$ podlega rozkładowi Poissona o parametrze $n \cdot \mu$

$$P(Z = z) = e^{-n\mu}\frac{(n\mu)^z}{z!} = Poi(z; -n\ln P) \qquad z = 0, 1, \ldots, \infty$$

Jest to rozkład dyskretny zatem rozkład estymatora $\hat{P}_2 = e^{-Z/n}$ też jest rozkładem dyskretnym i ma postać:

$$P(\hat{P}_2 = p_2 = e^{-\frac{z}{n}}) = Poi(-n\ln p_2; -n\ln P) \qquad p_2 = 0, \ldots, e^{-2/n}, e^{-1/n}, 1$$

Powyższy rozkład dla $n = 30$ i $P = e^{-2}$ odpowiada dolnemu wykresowi na rysunku 18.

Znając rozkład estymatora można obliczyć jego wartość oczekiwaną i tym samym otrzymać pierwszą z wielkości używanych do oceny estymatorów.

Definicja 15.6 (Estymator nieobciążony)**.**
Estymator $\hat{\Theta}$ parametru Θ jest nieobciążony jeśli

$$E[\hat{\Theta}] = \Theta$$

Różnicę $E[\hat{\Theta}] - \Theta$ nazywamy obciążeniem estymatora.

Przykład 15.4 (Obciążenie estymatora)**.**
Jakie są obciążenia estymatorów $\hat{P}_1, \hat{P}_2$?

$$E[\hat{P}_1] = E\left[\frac{k_0}{n}\right] = \frac{E[k_0]}{n} = \frac{nP}{n} = P$$

Estymator $\hat{P}_1$ jest estymatorem nieobciążonym. W przypadku estymatora $\hat{P}_2$ nie można skorzystać z liniowości wartości oczekiwanej ponieważ związek między estymatorem $\hat{P}_2$ a zmienna losową Z nie jest liniowy.

$$\begin{aligned} E[\hat{P}_2] &= \sum_{z=0}^{z=\infty} e^{-z/n} P^n \frac{(-n \ln P)^z}{z!} \\ &= P^n \sum_{z=0}^{z=\infty} \frac{\left(-e^{-1/n} n \ln P\right)^z}{z!} \\ &= P^n e^{-e^{-1/n} n \ln P} \\ &= P^{n(1-e^{-1/n})} \end{aligned}$$

gdzie skorzystaliśmy z rozwinięcia $e^x = \sum \frac{x^k}{k!}$.
Estymator $\hat{P}_2$ jest estymatorem obciążonym, jego obciążenie wynosi

$$E[\hat{P}_2] - P = P^{n(1-e^{-1/n})} - P$$

Dla $n = 30$ i $P = e^{-2}$ obciążenie estymatora wynosi $0,0045$.
Dla $n \to \infty$ wyrażenie $n(1 - e^{-1/n}) \to 1$ i tym samym

$$E[\hat{P}_2] \to P$$

Estymatory takie nazywamy asymptotycznie nieobciążonymi

Definicja 15.7 (Estymator asymptotycznie nieobciążony)**.**
Estymator $\hat{\Theta}$ parametru Θ jest asymptotycznie nieobciążony jeśli

$$\lim_{n\to\infty} E[\hat{\Theta}] = \Theta$$

Twierdzenie 15.1 (Nieobciążony estymator wartości oczekiwanej i wariancji)**.**
Załóżmy że próba losowa $X_1, X_2, \ldots, X_n$ *jest próbą prostą z rozkładu o nieznanej wartości oczekiwanej* μ *oraz wariancji* σ^2*. Nieobciążonymi estymatorami tych wielkości są:*

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{\sigma^2} = S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Dowód.

$$E[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

Aby dowieść drugą część twierdzenia, rozpatrzmy najpierw zmienną losową $Y_i = X_i - \bar{X}$. Na mocy liniowości wartości oczekiwanej oraz pierwszej części twierdzenia:

$$E[Y_i] = E[X_i - \bar{X}] = \mu - \mu = 0$$

Zatem

$$\text{Var}\,[Y_i] = E[Y_i^2] - (E[Y_i])^2 = E[Y_i^2]$$

$\square$

$$\begin{aligned}
E[\hat{\sigma^2}] &= \frac{1}{n-1}\sum_{i=1}^{n} E[(X_i - \bar{X})^2] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \text{Var}\,[X_i - \bar{X}] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \text{Var}\,[X_i - \frac{1}{n}\sum_{j=1}^{n} X_j] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \text{Var}\,[X_i - \frac{1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \text{Var}\left[\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \left[\frac{(n-1)^2}{n^2}\text{Var}\,[X_i] + \frac{1}{n^2}\sum_{j\neq i}\text{Var}\,[X_j]\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \left[\frac{(n-1)^2}{n^2}\sigma^2 + \frac{1}{n^2}(n-1)\sigma^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \frac{(n-1)}{n}\sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2
\end{aligned}$$

Czynnik $\frac{1}{n-1}$ w drugiej części twierdzenia 17.1 pojawia się dlatego, że nie znamy wartości oczekiwanej μ i zamiast niej używamy estymatora $\bar{X}$.

Twierdzenie 15.2 (Nieobciążony estymator wariancji)**.**
Załóżmy że próba losowa $X_1, X_2, \ldots, X_n$ jest próbą prostą z rozkładu o znanej wartości oczekiwanej μ oraz nieznanej wariancji σ^2. Nieobciążonym estymatorem wariancji będzie:

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Dowód.

$$\begin{aligned} E[\hat{\sigma^2}] &= \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \mu)^2] \\ &= \frac{1}{n}\sum_{i=1}^{n} \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2 \end{aligned}$$

□

Twierdzenie 17.1 uzupełnione estymatorami momentów wyższych rzędów stonowi podstawę ogólnej metody znajdowania estymatorów zwanej metodą momentów.

Definicja 15.8 (Moment rzędu n)**.**
Niech X jest zmienną losową. Momentem zmiennej X rzędu n nazywamy

$$\mu_n = E[X^n]$$

Definicja 15.9 (Estymator momentu rzędu n)**.**
Niech $X_1, X_2, \ldots, X_k$ jest próbą prostą. Estymatorem momentu (lub momentem z próby losowej) rzędu n nazywamy

$$\hat{\mu}_n = \frac{1}{k}\sum_{i=1}^{k} X_i^n$$

Twierdzenie 15.3 (Estymatory momentów)**.**
Niech $X_1, X_2, \ldots, X_k$ jest próbą prostą.

$$E[\hat{\mu}_n] = \mu_n; \quad \mathit{Var}[\hat{\mu}_n] = \frac{1}{k}\mathit{Var}[X^n]$$

Dowód.

$$\begin{aligned} E[\hat{\mu}_n] &= \frac{1}{k}\sum_{i=1}^{k} E[X_i^n] \\ &= \frac{k\mu_n}{k} = \mu_n \\ \mathrm{Var}[\hat{\mu}_n] &= \frac{1}{k^2}\sum_{i=1}^{k} \mathrm{Var}[X_i^n] \\ &= \frac{k\mathrm{Var}[X^n]}{k^2} = \frac{\mathrm{Var}[X^n]}{k} \end{aligned}$$

□

Definicja 15.10 (Metoda momentów)**.**
Niech nieznany parametr Θ rozkładu zm. los. X da się wyrazić jako funkcja pewnej liczby momentów tego rozkładu:

$$\Theta = f(\mu, \sigma^2, \mu_2, \ldots)$$

wtedy

$$\hat{\Theta} = f(\hat{\mu}, \hat{\sigma^2}, \hat{\mu_2}, \ldots)$$

Estymator $\hat{\Theta}$ nazywamy estymatorem metody momentów.

- Jeśli funkcja f jest liniowa to nieobciążoność estymatorów momentów przenosi się na estymator $\hat{\Theta}$ w przeciwnym przypadku estymator będzie obciążony.

Przykład 15.5 (Odległość punktów od środka okręgu)**.**
Dysponujemy próbą prostą $x_1, x_2, \dots x_n$ odległości losowo wybranych punktów wewnątrz okręgu o promieniu R od środka tego okręgu.
Jak skonstruować nieobciążony estymator nieznanej wartości promienia R ?

$$f(x; R) = \frac{2x}{R^2} \qquad 0 < x < R$$
$$\mu = \frac{2}{3}R$$

Nieobciążonym estymatorem R metody momentów jest

$$\hat{R}_1 = \frac{3}{2}\hat{\mu} = \frac{3}{2}\bar{X}$$

Wartości ekstremalne jako estymatory
Niech zm. los. X będzie określona na przedziale $[\alpha, \beta]$. Dysponując próbą prostą $x_1, x_2, \dots x_n$:

- $\hat{\alpha} = \text{Min }(x_1, x_2, \dots, x_n)$
- $\hat{\beta} = \text{Max }(x_1, x_2, \dots, x_n)$

Przykład 15.6 (Odległość punktów od środka okręgu)**.**
Czy $\hat{\beta} = \text{Max }(x_1, x_2, \dots, x_n)$ jest nieobciążonym estymatorem parametru R z poprzedniego przykładu?

Twierdzenie 15.4 (Rozkład wartości maksymalnej)**.**
Niech $X_1, X_2, \dots, X_n$ będą niezależnymi zm. los. o dystrybuancie F_X i gęstości pr. f_X oraz niech $Z = \text{Max }(X_1, X_2, \dots, X_n)$. Wtedy

- $F_Z(z) = (F_X(z))^n$
- $f_Z(z) = \frac{d}{dz}\left[(F_X(z))^n\right] = n(F_X(z))^{n-1}f_X(z)$

Dowód.

$$\begin{aligned} F_Z(z) &= P(Z \leqslant z) \\ &= P(\text{Max }(X_1, X_2, \dots, X_n) \leqslant z) \\ &= P(X_1 \leqslant z, X_2 \leqslant z, \dots, X_n \leqslant z) \\ &= P(X_1 \leqslant z) \cdot P(X_2 \leqslant z) \cdots P(X_n \leqslant z) \\ &= (F_X(z))^n \end{aligned}$$

□

Twierdzenie 15.5 (Rozkład wartości minimalnej)**.**
Niech $X_1, X_2, \dots, X_n$ będą niezależnymi zm. los. o dystrybuancie F_X i gęstości pr. f_X oraz niech $V = \text{Min }(X_1, X_2, \dots, X_n)$. Wtedy

- $F_V(v) = 1 - [1 - F_X(v)]^n$

- $f_V(v) = \dfrac{d}{dv}\left[1 - [1 - F_X(v)]^n\right] = n[1 - F_X(v)]^{n-1} f_X(v)$

Dowód.

$$\begin{aligned}
F_V(v) &= P(V \leqslant v) \\
&= P(\text{Min }(X_1, X_2, \ldots, X_n) \leqslant v) \\
&= 1 - P(\text{Min }(X_1, X_2, \ldots, X_n) > v) \\
&= 1 - P(X_1 > v, X_2 > v, \ldots, X_n > v) \\
&= 1 - [1 - P(X_1 \leqslant v)] \cdot [1 - P(X_2 \leqslant v)] \cdots [1 - P(X_n \leqslant v)] \\
&= 1 - [1 - F_X(v)]^n
\end{aligned}$$

□

Przykład 15.7 (Odległość punktów od środka okręgu)**.**
Czy $Z = \text{Max }(X_1, X_2, \ldots, X_n)$ jest nieobciążonym estymatorem parametru R z poprzedniego przykładu?

Rozwiązanie.

$$F_X(x) = \left[\frac{x}{R}\right]^2 \qquad 0 < x < R$$

$$f_Z(z) = \frac{2n}{R}\left[\frac{z}{R}\right]^{2n-1} \qquad 0 < z < R$$

$$E[Z] = \frac{2n}{R^{2n}} \int_0^R z^{2n} dz = \frac{2n}{2n+1} R$$

Zatem zmienna losowa Z jest estymatorem obciążonym ale

$$\hat{R}_2 = \frac{2n+1}{2n} Z; \qquad E[\hat{R}_2] = R$$

jest już estymatorem nieobciążonym.

Definicja 15.11 (Efektywność estymatora)**.**
Niech $\hat{\Theta}_1$ i $\hat{\Theta}_2$ będą nieobciążonymi estymatorami tego samego parametru Θ. Estymator $\hat{\Theta}_1$ nazywamy bardziej efektywnym od estymatora $\hat{\Theta}_2$ wtedy gdy

$$\text{Var }[\hat{\Theta}_1] < \text{Var }[\hat{\Theta}_2]$$

Przykład 15.8 (Efektywność estymatora)**.**
Który z estymatorów $\hat{R}_1$ czy $\hat{R}_2$ jest bardziej efektywny?

$$f(x; R) = \frac{2x}{R^2}; \qquad \text{Var }[X] = \frac{R^2}{18}$$

$$\begin{aligned}
\text{Var }[\hat{R}_1] &= \text{Var }\left[\frac{3}{2}\bar{X}\right] \\
&= \text{Var }\left[\frac{3}{2}\frac{\sum_{i=1}^{n} X_i}{n}\right] \\
&= \frac{9}{4n^2}\sum_{i=1}^{n} \text{Var } X_i \\
&= \frac{9}{4n^2} n \frac{R^2}{18} \\
&= \frac{R^2}{8n}
\end{aligned}$$

$$Z = \text{Max}\,(X_1, X_2, \ldots, X_n)$$
$$E[Z] = \frac{2n}{2n+1}R$$
$$E[Z^2] = \frac{2n}{R^{2n}}\int_0^R z^{2n+1}dz = \frac{2n}{2n+2}R^2 = \frac{n}{n+1}R^2$$

$$\begin{aligned}
\text{Var}\,[Z] &= E[Z^2] - E[Z]^2 \\
&= \frac{n}{n+1}R^2 - \frac{4n^2}{(2n+1)^2}R^2 \\
&= \frac{n(2n+1)^2 - 4n^2(n+1)}{(n+1)(2n+1)^2}R^2 \\
&= \frac{n}{(n+1)(2n+1)^2}R^2 \\
\text{Var}\,[\hat{R}_2] &= \text{Var}\left[\frac{2n+1}{2n}Z\right] \\
&= \frac{(2n+1)^2}{4n^2}\frac{n}{(n+1)(2n+1)^2}R^2 \\
&= \frac{R^2}{4n(n+1)}
\end{aligned}$$

$$\text{Var}\,[\hat{R}_1] = \frac{R^2}{8n} \quad > \quad \text{Var}\,[\hat{R}_2] = \frac{R^2}{4n(n+1)}$$

Estymator $\hat{R}_2$ jest bardziej efektywny niż estymator $\hat{R}_1$.

$$\lim_{n\to\infty} \text{Var}\,[\hat{R}_1] = \lim_{n\to\infty} \text{Var}\,[\hat{R}_2] = 0$$

Definicja 15.12 (Estymator zgodny)**.**
Estymator $\hat{\Theta}$ parametru Θ jest zgodny jeśli

$$\lim_{n\to\infty} \text{Var}\,[\hat{\Theta}] = 0$$

W pewnych dość często spotykanych problemach możemy określić dolne ograniczenie na wariancję nieobciążonego estymatora.

Twierdzenie 15.6 (Twierdzenie Cramera-Rao)**.**
Niech $X_1, X_2, \ldots, X_n$ jest próbą prostą o gęstości $f(x;\Theta)$, Takiej, że:

- $\forall x, \quad \exists\ \frac{\partial}{\partial\Theta}\ln f(x;\Theta)$
- *jeśli X jest ograniczona to ograniczenie nie zależy od Θ*
- *jeśli X jest nieograniczona to $f(x;\Theta)$ ma ciągłą pochodną i jest całkowalna dla każdej wartości Θ.*

Wtedy wariancja nieobciążonego estymatora $\hat{\Theta}$ spełnia nierówność:

$$\mathit{Var}\,[\hat{\Theta}] \geqslant \frac{1}{nE[(\frac{\partial}{\partial\Theta}\ln f(x;\Theta))^2]}$$

Definicja 15.13 (Estymator najefektywniejszy)**.**
Estymator nieobciążony którego wariancja przyjmuje wartość minimalną

$$\text{Var}\,[\hat{\Theta}] = \frac{1}{nE[(\frac{\partial}{\partial\Theta}\ln f(x;\Theta))^2]}$$

nazywamy estymatorem najefektywniejszym.

W przypadku estymatorów obciążonych podstawą oceny jakości estymatora nie może być tylko jego wariancja. W takich wypadkach podstawą oceny jakości estymatora jest średni błąd kwadratowy.

Definicja 15.14 (Średni błąd kwadratowy estymatora)**.**
Niech $\hat{\Theta}$ będzie estymatorem parametru Θ. Średnim błędem kwadratowych estymatora nazywamy:

$$\begin{aligned}
\text{MSE}\,[\hat{\Theta}] &= \text{E}\,[(\hat{\Theta}-\Theta)^2] \\
&= \text{E}\,[(\hat{\Theta}-E[\hat{\Theta}]+E[\hat{\Theta}]-\Theta)^2] \\
&= \text{E}\,[(\hat{\Theta}-E[\hat{\Theta}])^2+2(\hat{\Theta}-E[\hat{\Theta}])(E[\hat{\Theta}]-\Theta) \\
&\quad +(E[\hat{\Theta}]-\Theta)^2] \\
&= \text{E}\,[(\hat{\Theta}-E[\hat{\Theta}])^2]+2(E[\hat{\Theta}]-\Theta)E[\hat{\Theta}-E[\hat{\Theta}]] \\
&\quad +E[(E[\hat{\Theta}]-\Theta)^2] \\
&= \text{Var}\,[\hat{\Theta}]+(E[\hat{\Theta}]-\Theta)^2 \\
&= \text{Wariancja}+\text{Obciążenie}^2
\end{aligned}$$

Średni błąd kwadratowy który jest wartością oczekiwaną kwadratu różnicy między estymatorem a wartością estymowaną jest sumą wariancji estymatora i kwadratu jego obciążenia. Uwaga: Estymatory obciążone mogą mieć mniejszy średni błąd kwadratowy niż dolne ograniczenie na wariancję estymatora nieobciążonego dane przez twierdzenia Cramera-Rao.
Nie zawsze da się w prosty sposób zaproponować estymator wielkości którą jesteśmy zainteresowani. Istnieje ogólna metoda wyznaczania estymatorów.

Definicja 15.15 (Zasada największej wiarygodności)**.**
Dysponując wynikami eksperymentu losowego $x_1, x_2, \ldots, x_n$ należy tak dobrać estymowane parametry aby wyniki te były najbardziej wiarygodne (bardziej prawdopodobne niż te same wyniki przy innych wartościach estymowanych parametrów).

Definicja 15.16 (Funkcja wiarygodności)**.**
Niech $\vec{x} = x_1, x_2, \ldots, x_n$ jest realizacją próby losowej $\vec{X} = X_1, X_2, \ldots, X_n$ o rozkładach zależnych od parametru Θ. Zasada największej wiarygodności sugeruje estymację parametru Θ wartością która maksymalizuje funkcję będącą łącznym pr. (gęstością pr.) realizacji próby losowej rozumianej jako funkcję parametru Θ. Funkcję tę nazywamy funkcją wiarygodności

$$L(\Theta) = L(\Theta;\vec{x}) = P(\vec{X}=\vec{x};\Theta) = P(\vec{X})$$

$$L(\Theta;\vec{x}) = f_{\vec{X}}(\vec{x};\Theta)$$

Funkcja wiarygodności
Jeśli $\vec{X} = X_1, X_2, \ldots, X_n$ jest próbą prostą:

$$L(\Theta) = p(x_1;\Theta)p(x_2;\Theta)\cdots p(x_n;\Theta)$$

$$L(\Theta) = f(x_1;\Theta)f(x_2;\Theta)\cdots f(x_n;\Theta)$$

Jeśli próba nie jest prosta to:

$$L(\Theta) = p_1(x_1;\Theta)p_2(x_2|x_1;\Theta)\cdots p_n(x_n|x_1,x_2,...,x_{n-1};\Theta)$$

$$L(\Theta) = f_1(x_1;\Theta)f_2(x_2|x_1;\Theta)\cdots f_n(x_n|x_1,x_2,...,x_{n-1};\Theta)$$

Definicja 15.17 (Estymator metody największej wiarygodności)**.**
Estymatą metody największej wiarygodności jest wartość $\theta = h(x_1, x_2, \ldots, x_n)$ która maksymalizuje (zwykle zeruje pierwszą pochodną) funkcję wiarygodności $L(\Theta)$. Estymata ta jest realizacją estymatora

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

zwanego estymatorem metody największej wiarygodności.

- $L(\hat{\Theta}) = \text{Sup}\ \{L(\Theta)\}; \quad \Rightarrow \frac{\partial L(\Theta)}{\partial \Theta}|_{\Theta=\hat{\Theta}} = 0$
- $\hat{\text{Var}}\ [\hat{\Theta}] \approx \frac{-1}{\frac{\partial^2 ln L(\Theta)}{\partial \Theta^2}|_{\Theta=\hat{\Theta}}}$

Postać estymatora wariancji estymatora metody największej wiarygodności wynika z faktu, że dla dużych próbek estymator ma w przybliżeniu rozkład normalny.
Estymatory metody największej wiarygodności są:

- niezmiennicze - Jeśli $\hat{\Theta}$ jest estymatorem największej wiarygodności parametru Θ to $g(\hat{\Theta})$ jest estymatorem największej wiarygodności wielkości $g(\Theta)$, (jeśli g jest odwracalna).
- asymptotycznie nieobciążone;
- asymptotycznie najefektywniejsze -

$$\hat{\text{Var}}\ [\hat{\Theta}] \rightarrow \frac{1}{nE[(\frac{\partial}{\partial \Theta} \ln f(x))^2]}$$

Ostatnia własność oznacza, że estymator wariancji estymatora dąży do dolnego ograniczenia danego przez twierdzenie Cramera-Rao.