Wykład 1, *04.10.2010*

1. **Lemat o minimaxie** - $\sup_\pi \inf_\delta L(\pi,\delta) \leqslant \inf \delta \sup_\pi L(\pi,\delta)$, gdzie $L : \Pi \times \Delta \to \mathbb{R}$.
2. **Gra ma wartość**, jeśli $\sup_\pi \inf_\delta L(\pi,\delta) = \inf_\delta \sup_\pi L(\pi,\delta)$.
3. δ^* jest **minimaxowa**, jeśli $\sup_\pi L(\pi,\delta^*) = \inf_\delta \sup_\pi L(\pi,\delta)$.
4. π^* jest **maximinowe**, jeśli $\inf_\delta L(\pi*,\delta) = \sup_\pi \inf_\delta L(\pi,\delta)$.
5. δ^* jest π^***-optymalna** (π^*-bayesowska), jeśli $L(\pi^*,\delta^*) = \inf_\delta L(\pi^*,\delta)$.
6. (π^*,δ^*) jest **siodłem** (punktem siodłowym, in. punktem równowagi Nasha) gry, jeśli $\forall \pi,\delta\ L(\pi,\delta^*) \leqslant L(\pi^*,\delta^*) \leqslant L(\pi^*,\delta)$, in. $\sup_\pi L(\pi,\delta^*) = L(\pi^*,\delta^*) = \inf_\delta(\pi^*,\delta)$.
7. **Tw. o siodle** - Ustalmy $\pi^*,\delta^*)$. N.w.s.r.:
   (a) $\sup_\pi L(\pi,\delta^*) \leqslant \inf_\delta L(\pi^*,\delta)$
   (b) (π^*,δ^*) jest siodłem
   (c) (π^*,δ^*) jest siodłem i δ^* jest π^*-optymalne
   (d) π^* jest maximinowe, δ^* jest minimaxowa i gra ma wartość
   (e) π^* jest maximinowe, δ^* jest minimaxowa i gra ma wartość $= L(\pi^*,\delta^*)$
8. **Wniosek** - Jeżeli δ^* jest t. że ma stałe ryzyko tzn. $\forall \pi_1,\pi_2\ L(\pi_1,\delta^*) = L(\pi_2,\delta^*)$ oraz $\exists \pi*$ t. że δ^* jest π^*-optymalna, to (π^*,δ^*) jest siodłem.
9. H - zb. czystych strategii natury ($\theta \in H$ - nieznany parametr rozkładu prawdopodbieństwastwa), A - zb. czystych strategii statystyka ($a \in A$ - akcje statystyka). Zrandomizowane strategie natury (zbiór $\Pi = H^*$); zrandomizowane strategie statystyka ($\Delta = A^*$); funkcja straty z przestrzeni zrandomizowanych.
10. **Def.** $L(\pi,\delta) = E_{\pi\times\delta} L(\theta,a)$.
11. **Twierdzenie von Neumanna** - Jeżeli mamy grę (H,A,L) ze skończonymi zb. H i A, to gra (H^*,A^*,L) ma punkt siodłowy.

Wykład 2, *11.10.2010*

12. **Oznaczenia**: H - zbiór czystych strategii natury, A - zbiór akcji statystyka, X - przestrzeń obserwacji (polska), $\{P_\theta : \theta \in H\}$ - rozkłady prawdopodobieństwa na X, $X \sim P_\theta$, $L : H \times A \to \mathbb{R}$ - funkcja straty, $\Pi = \{\pi :$ rozkład prawdopodobieństwa na H$\}=H^*$, D - zbiór reguł decyzyjnych, D^* - rodzina rozkładów prawdopodobieństwa na zbiorze wszystich reguł decyzyjnych (niezrandomizowanych).
13. **Reguła decyzyjna** (czysta strategia statystyka, np. estymator, test) $d : x \in X \to d(x) \in A$.
14. **Ryzyko:** $R(\theta, d) = E_\theta L(\theta, d(X)) = \int_X L(\theta, d(x)) P_\theta(dx)$.
15. Gra (H, D, R) zamiast (H, A, L), gdzie $D = \{d : X \to A$ mierzalne, takie że ryzyko $R(\theta, d)$ jest dobrze zdefiniowane$\}$; $R : H \times D \to \mathbb{R}$.
16. **Ryzyko bayesowskie** $f(\pi, d) = \int_H R(\theta, d)\pi(d\theta)$.
17. Gra: niezdrandomizowana (H, D, R), gra zrandomizowana $(H^*$, D^*, $R^*)$, d^* - zrandomizowana reguła decyzyjna.
18. $R^*(\pi, d^*) = \int_D r(\pi, d) d^*(dd)$
19. $\delta : X \to A^*$ - zbiór rozkładów prawdopodobieństwa na A, $x \in X \to \delta(x)$ - miara probabilistyczna na zbiorze akcji A.
    $r(\pi, \delta) = \int_X \int_X \int_A L(\theta, a)\delta(x, da) P_\theta(dx)\pi(d\theta)$.
20. **Tw. Walda-Wolfowitza** Dla każdej reguły zrandomizowanej $d^* \in D^*$ istnieje reguła behawiorystyczna $\delta \in \tilde{D}$ taka że $R^*(\pi, d^* = r(\pi, \delta)$.
21. **Zbiór ryzyka - dwie hipotezy proste** $\{R(\cdot, d),\ d \in D\} =\}(\alpha(d),\ \beta(d))$, $d \in D\}$, gdzie $\alpha(d) = R(0, d) = \int_{\{x:d(x)=1\}} p_0(x)dx$ - błąd I rodzaju, $\beta(d) = R(1, d) = \int_{\{x:d(x)=0\}} p_1(x)dx$ - bład drugiego rodzaju. Albo też: $\alpha(\delta) = \int_X \delta(x)p_0(x)dx$, $\beta(\delta) = \int_X (1 - \delta(x))p_1(x)dx$.

Wykład 3, *18.10.2010*

22. **Oznaczenia c.d.:** d - przestrzeń niezrandomizowanych reguł, Δ - przestrzeń zrandomizowanych (behawiorystycznych) reguł decyzyjnych $x \in X \rightarrow \delta(x)$ - miara probabilistyczna na A.
$R : H \times \Delta \rightarrow \mathbb{R}$, $R(\theta, \delta) = \int_X \int_A L(\theta, a)\delta(x, da)P_\theta(dx)$.

23. **Def.** $\delta_1 \preceq \delta_2 \Leftrightarrow \forall_\theta \ R(\theta, \delta_1) \leqslant R(\theta, \delta_2)$, $\delta_1 \prec \delta_2 \Leftrightarrow \delta_1 \preceq \delta_2$ i $\exists_\theta \ R(\theta, \delta_1) < R(\theta, \delta_2)$, $\delta_1 \sim \delta_2 \Leftrightarrow \delta_1 \preceq \delta_2$ i $\delta_2 \preceq \delta_1$.

24. **Def.** $C \subseteq \Delta$, mówimy, że C jest klasą **zupełną**, jeśli $\forall \delta \notin C \ \exists_{\delta' \in C}$ taka że $\delta' \prec \delta$. C jest klasą **istotnie zupełną**, jeśli $\forall_\delta \ \exists_{\delta' \in C}$ taka że $\delta' \preceq \delta$.

25. **Def.** Reguła δ jest **dopuszczalna**, jeśli nie istnieje reguła ściśle lepsza od niej tzn. nie istnieja δ' taka że $\delta' \prec \delta$. Oznaczenie: Adm - rodzina reguł dopuszczalnych.

26. **Lemat** jeśli C jest zupełna, to $Adm \subseteq C$.

27. **Lemat** Jeśli C jest istotnie zupełna, $\delta \in Adm \backslash C$, to $\exists \delta' \in C$, $\delta' \sim \delta$.

28. **Def. Zbiór ryzyk:** $R = \{R(\delta) : \delta \in \Delta\} \subseteq \mathbb{R}^k$.

29. **Stw.** R jest wypukły, domknięty i ograniczony.

30. **Def. Reguła δ jest bayesowska** względem rozkładu a priori π na H, jeśli $r(\pi, \delta) = \inf_\delta R(\pi, \delta)$.

31. **Stw.** Jeśli reguła δ jest bayesowska względem π, $(\pi_i > 0 \ \forall i)$, to $\delta \in Adm$.

32. **Stw.** Jeżeli reguła δ jest dopuszczlna,a to jest bayesowska względem pewnego rozkładu a priori π.

33. **Def.** Jeżeli $Z \subset \mathbb{R}^k$ jest wypukły, domknięty, ograniczony, to dolny brzeg tego zbioru określamy $\lambda(Z) := \{\alpha : Z \cap Q_\alpha = \{a\}\}$.

34. **Uwaga** $\lambda(Z) \subseteq \delta(Z)$.

35. **Lemat** Każdy domknięty, wypukły, ograniczony zbiór Z ma $\lambda(Z) \neq \emptyset$.

Wykład 4, *25.10.2010*

36. **Tw. Rao-Blackwella** Mamy zbiór $A \in \mathbb{R}^d$. $L(\theta, a)$ - wypukła funkcja a dla każdego $\theta \in \Theta$. Niech T będzie statystyką dosteczną dla rodziny P_θ. $d \in D$ (niezrandomizowana decyzja) $\Rightarrow d_0(x) = E(d(X)|T = t)$ jest nie gorsza niż d.

37. **Def. Statystyka dostateczna** - $T : X \to \mathbb{R}^d$. $\forall A \in B(X)$ $P_\theta(X \in A|T = t)$ - nie zależy od $\theta \Leftrightarrow f_\theta(x) = g_\theta(T(X))h(X)$ - kryterium faktoryzacji, g_θ - zależy od θ przez statystykę.

38. **Twierdzenie** Mamy $A = (-\infty, \infty)$, $\forall\theta \in \Theta$ $L(\theta, a)$ jest wypukła. $\exists\theta_0$ taka że $L(\theta_0$, a)$\to \infty$, $|a| \to \infty$, to klasa reguł niezrandomizowanych jest istotnie zupełna dla (Θ, D, R).

39. **Uwaga** $L(\theta_0, a) \geqslant c|a| + b$ $\forall a \in A$.

40. **Twierdzenie o rozdzielaniu zbiorów wypukłych** Jeśli $U, W \in \mathbb{R}^d$ są wypukłe i $intU \cap intW = \emptyset$, to $\exists \bar{\pi} \in \mathbb{R}^d$, $\pi \neq 0$ takie że $\pi(u) \leqslant c \leqslant \pi(w)$ $\forall u \in U$ $\forall w \in W$ $\{u : \pi(u) = c\}$.

Wykład 5, *08.11.2010*

41. **Rodziny wykładnicze** $f_\theta(x) = exp\{< \phi(\theta), T(x) > -A(\theta)\}h(x)$, gdzie $\theta \in \Theta \subseteq \mathbb{R}^k$.
    Naturalna parametryzacja rodziny wykładniczej.: $\eta = \phi(\theta)$, $f_\eta(x) = exp\{< \eta, T(x) > -A(\eta)\}h(x)$.

42. **Przykład**
    Weźmy $X_1, \ldots, X_n$ obserwacji, $x|\theta \sim f_\theta(x)$, $f_\theta(x) = exp\{< \theta, T(x) - A(\theta)\}h(x)$ - **rozkład obserwacji**
    $\theta|\mu, \nu$, $\nu \in \mathbb{R}^k$, $\nu \in \mathbb{R}$ - k+1 parametrów, $\Pi(\theta) = exp\{< \theta, \mu > -\nu A(\theta)\}$ - **rozkład a priori**
    $f_\theta(x_1, \ldots, x_n) \sim exp\{< \theta, \sum T(x_i > -nA(\theta)\}$ - **rozkład łączny**
    $\Pi(\theta|x, \mu, \nu) = f_\theta(x_1, \ldots, x_n)\Pi(\theta|\mu, \nu) = exp\{< \theta, \mu + \sum T(x_i) > -(n + \nu)A(\theta)\}$ - **rozkład a posteriori**
    Rozdkłady a priori i a posteriori są **sprzężone**.

43. **Funkcje tworzące momenty** ${}^mT(x)^{(t)} = Ee^{<t,T(x)>}$.

Wykład 6, *15.11.2010*

44. Model hierarchiczny (bayesowski). Łańcuchy Markowa: algorytm Metropolisa - Hastingsa, próbnik Gibbsa.

Wykład 7, *22.11.2010*

45. **Gra w kodowanie**.
46. **Podstawy statystyki bayesowskiej** - konstrukcja modelu bayesowskiego:
X - przestrzeń obserwacji
$\{P_\theta : \theta \in \Theta\}$ - rodzina rozkładów prawdopodobieństwa na X (niewiele tracimy od razu posługując się gęstościami)
μ - miara na X (mamy σ ciało)
f_θ - gęstość P_θ względem μ; $P_\theta(A) = \int_A f_\theta(x)\mu(dx)$
Θ - przestrzeń parametrów π - rozkład prawdopodobieństwa na Θ
ν - miara na Θ
π - gęstość π względem ν (a priori).
47. **Def.** $f(\theta, x) = (df)\pi(\theta)f_\theta(x)$ - łączna gęstość na $\Theta \times X$ względem miary $\nu \times \mu$.
48. $f(x|\theta) = \frac{f(\theta,x)}{\pi(\theta)} = f_\theta(x)$
Wzór Bayesa, rozkład a posteriori ($\pi_x(\theta)$): $f(\theta|x) = \frac{f(x|\theta)\cdot\pi(\theta)}{f(x)}$ - rozkład łączny w liczniku; gdzie $f(x) = \int_\Theta f_\theta(x)\pi(\theta)\nu(d\theta)$.
49. **Model statystyczny** - podanie rodziny rozkładów prawdopodobieństwa zależnych od θ. **Model bayesowski** - rozkład łączny, dwuwymiarowy.
50. $L(\theta, a)$ - argumenty: parametr i akcja, ktorą podejmujemy, $L : \Theta \times A \to \mathbb{R}$, gdzie A - zbiór akcji. **Reguła zrandomizowana** $\delta : X \to A$.
51. Reguła jest π^***-bayesowska**, jeśli minimalizuje ryzyko bayesowskie, tzn. jeśli $r(\pi, \delta_\pi) \inf_\delta r(\pi, \delta)$.
52. **R - ryzyko niebayesowskie**
$R(\theta, \delta) = E_\theta L(\theta, \delta(x)) = \int_X L(\theta, \delta(x))f_\theta(x)\mu(dx) = E[L(\theta, \delta(x))|\nu = \theta]$.
53. **Ryzyko bayesowskie** $r(\pi, \delta) = EL(\nu, \delta(x)) = \int_\Theta R(\theta, \delta)\pi(\theta)\nu(d\theta) = \int_\Theta \int_X L(\theta, \delta(x))f_\theta(x)\mu(dx)\pi(\theta)\nu(d\theta) = (*)$.
54. **Podstawowe tw. statystyki bayesowskiej** Jeżeli $\forall x\ \exists a^* = \delta^*(x)$ taka że $E[L(\nu, \delta^*(x))|X = x] = \inf_{a\in A} E(L(\delta, a)|X = x)$ to δ^* jest regułą bayesowską. $E[L(\nu, a)|X = x] = \int_\Theta L(\theta, a)\pi_x(\theta)d\theta$ - w zadaniach chcemy minimalizować tą całkę (po skorzystaniu z tw. Fubiniego mamy, że $(*) = r(\pi, \delta) = \int_X \int_\Theta L(\theta, \delta(x))\pi_x(\theta)d\theta f(x)dx)$.
55. **Optymalna reguła a priori** $a^* = argmin \int_\Theta L(\theta, a)\pi(\theta)d\theta$.

WYKŁAD 8, *29.11.2010*

56. **Teorio-decyzyjne podstawy analizy dyskryminacyjnej**: obserwujemy obiekt należący do jednej z podpopulacji $1, \ldots, k$ i mamy zdecydować, do której należy na podstawie obserwowanych cech $X = (X_1, \ldots, X_d)$, np. diagnoza medyczna, bankowość, rozpoznawanie symboli.

57. Matematycznie:
X - wektor losowy w $\mathbb{R}^d$ X ma w podpopulacji i rozkład f_i (gęstość)
π_i - prawdopodobieńtwo a priori, że obiekt należy do i-tej podpopulacji $\pi_1 + \pi_k = 1$. Zakładamy, że $f_1, \ldots, f_k$ i $\pi_1, \ldots, \pi_k$ - znane.
$\Theta = \{1, \ldots, k\}$, $\theta = i$. Musimy wybrać, z którego rozkładu prawdopodobieństwa pochodzi obserwacja, z której podpopulacji: $A = \Theta$ lub $A = \Theta \cup \{0\}$ (0 - zawieszanie decyzji)
Funkcja strat - macierz $L(i, j)$, $i \in \Theta$, $j \in A$.

58. **Wzór Bayesa** (dyskretna wersja): $\pi(i|x) = \frac{f_i(x)\pi_i}{f(x)}$, $f(x) = \sum_{i=1}^k f_i(x)\pi_i$.

59. **Reguła klasyfikacyjna** $\delta : X \to A$.

60. **Ryzyko bayesowskie**
$r(\pi, \delta) = r(\delta) = EL(I, \delta(X)) = \sum_{i=1}^k \pi_i \int_X f_i(x)L(i, \delta(x))dx$.

61. **Reguła bayesowska** δ^* - minimalizująca ryzyko bayesowskie.

62. **Ryzyko a posteriori** $\sum_{i=1}^k \pi(i|x)L(i, j) \to \min j \in A$.

63. **Wniosek (podstawowa reguła bayesowska)**
$\delta^*(x) = arg \min_j \sum_{i=1}^k \pi(i|x)L(i, j)$.

64. **Klasyczny przykład** (wielowymiarowe rozkłady normalne): lda ($f_i = N(\mu_i, V)$), qda ($f_i = N(\mu_i, V_i)$=- dwie podstawowe funkcje.

65. **Obszar decyzyjny** $D_j = \{x : \delta^*(x) = j\}$.

66. **Estymacja nieznanych rozkładów w klasach** f_i.

Wykład 9, *06.12.2010*

67. **Modele liniowe** Predykcja: $X = (X_1, \ldots, X_n)$, Y - zmienna losowa. SSzukamy $Y = h(X) = c_0 + \sum_{i=1}^n c_i X_i$.

68. **Def.** $\hat{Y} = h(X)$ jest **najlepszym liniowym predyktorem Y (BLP(Y))**, jeśli:

    - $h : \mathbb{R}^N \to \mathbb{R}$ jest funkcją liniową
    - $E(Y - h(X))^2 \leqslant E(Y - g(X))^2$, $\forall g$: $\mathbb{R}^n \to \mathbb{R}$, liniowa

69. **Twierdzenie** $\hat{Y} = h(X) = c_0 + \sum_{i=1}^n c_i X_i$ jest BLP(Y), jeśli $(c_0, c_1, \ldots, c_n)$ spełniają układ równań:

$$\begin{cases} \sum_{i=1}^n Cov(X_i, X_k) = Cov(X_k, Y), \quad k = 1, \ldots, n \\ c_0 = EY - \sum_{i=1}^n c_i EX_i. \end{cases}$$

70. **Def.** $\hat{Y} = g(X)$ jest **nieobciążonym predyktorem Y**, jeśli $E_m(Y - g(X)) = 0 \ \forall m$.

71. **Def.** $\hat{Y} = BLUP(Y)$ **- best linear unbiased predictor**, jeśli:

    - $h(X)$ jest nieobciążony
    - $h : \mathbb{R}^n \to \mathbb{R}$ jest funkcją liniową
    - $E_m(Y - h(X))^2 \leqslant E_m(Y - g(X))^2$, $\forall g(X) = \tilde{Y}$ - nieobciążony, liniowy.

72. **Uwaga** $Y \to m$, to $BLUP \to BLUE$.

73. **Def.** $\hat{m} = g(X)$ nazywamy **nieobciążonym estymatorem m**, jeżeli $E_m g(X) = m$.

74. **Def.** $\hat{m} = BLUE(m)$, jeśli:

    - jest liniową funkcją
    - jest nieobciążony
    - $Var_m h(X) \leqslant Var_m g(X) \ \forall g(X) = \tilde{m}$ - nieobciążony, liniowy.

75. Od tej pory przyjmujemy założenia:

    (i) $E_m X_i = E_m Y = m$

    (ii) $Cov_m(X_i, X_k) = Cov(X_i, X_k)$

    (iii) $Cov_m(X_i, Y) = Cov(X_i, Y)$.

76. **Twierdzenie** Przy warunkach $(i)-(iii)$ $h(X) = c_0 + \sum_{i=1}^n c_i X_i$ jest BLUP(Y), jeżeli:
    (W1): $c_0 = 0$, $\sum_{i=1}^n c_i = 1$
    (W2): $\exists \lambda \in \mathbb{R}$ $\sum_{i=1}^n c_i Cov(X_i, X_k) - Cov(X_k, Y) = \lambda$, $k = 1, \ldots, n$.

77. **Twierdzenie** Przy warunkach $(i) - (iii)$, jeżeli $BLP(Y) = h(X, m)$ i jeżeli uda nam siępoliczyć $\hat{m} = BLUE(m)$, to $BLUP(Y) = h(X, \hat{m})$.

Wykład 10, *13.12.2010* **Predykcja/liniowa predykcja**

78. **Predyktory (od najlepszego do najgorszego):**

- **BP** - najlepszy predyktor wśród wszystkich funkcji X; $r^*(x) = E(Y|X)$ - wartość oczekiwana a posteriori w modelach bayesowskich. Trzeba znać rozkłady prawdopodobieństwa $f_\theta(x)$, $\pi(\theta)$, $E(\theta_j|X_{j1}, \ldots, X_{jn_j})$.
- **BLP** - best linear predictor, najlepszy predyktor wśród liniowych, $r^* = bx^* + a^*$. $b^* = \frac{COV(Y,X)}{VarX}$, $a^* = EY - b^*$. Dla $X = \mathbb{R}^n$ $r^*(x) = a^* + \sum_{i=1}^n b_i^* x_i$. Wówczas skalarnie:
  $\sum_{i=1}^n b_i^* Cov(X_i, X_k = Cov(X_k, Y)$, $k = 1, \ldots, n$,
  $a^* = EY - \sum_{i=1}^n b_i^* EX_i$.
  $BLP(X_{n+1}) = BLP(\mu(\theta)) = z\bar{X} + (1-z)m$, gdzie $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $z = \frac{na^2}{na^2+s^2}$ - współczynnik zaufania.
  Trzeba znać a^2, s^2, m.
- **BLUP** - trzeba znać komponenty wariancyjne (i BLUE - best linear unbiased estimator)
- **EBLUP** - nic nie trzeba znać, trzeba wierzyć w model.

79. **Model bayesowski** - podstawowy model statystyki bayesowskiej:
$(X_1, \ldots, X_n, X_{n+1}) \sim_{i.i.d.} f_\theta(\cdot)$
$\pi(\cdot)$ - gęstość a priori θ: $f(\theta, x_1, \ldots, x_n, x_{n+1}) = \pi(\theta)\pi_i f_\theta(x_i)$.
Dobrze jest operować skrótami: $u(\theta) = E(X_i|\theta) = \int x_i f_\theta(x_i) dx_i$.
$\sigma^2(\theta) = Var(X_i|\theta)$
$m = E\mu(\theta) = EX_i$
Komponenty wariancyjne, $VarX_i = s^2 + a^2$:
$a^2 = Var\mu(\theta)$
$s^2 = E\sigma^2(\theta) = \int \sigma^2(\theta)\pi(\theta)d\theta$.

WYKŁAD 11, *20.12.2010*

80. **Model Bühlmanna Strauba**:

- $\theta_1, X_{11}, \ldots, X_{1n_1}$
  $\ldots$
  $\theta_p, X_{p1}, \ldots, X_{pn_p}$
  gdzie X_{ij} - szkody j-tego klienta w i-tym roku, θ_j - zmienna strukturalna dla j-tego wiersza
- założenie bayesowskie, łączny rozkład: $f((\theta_j),(x_{ji})) = \prod_j \pi(\theta_j \prod_i f_{\theta_j}(x_{ji})$
- empiryczne podejście bayesowskie:
  $\mu(\theta_j = E(X_{ji}|\theta_j, \frac{\sigma^2(\theta_j)}{w_{ji}} = Var(X_{ji}|\theta_j)$, $m = EX_{ji} = E\mu(\theta)$, $s^2 = E\sigma^2(\theta_j)$, $a^2 = Var\mu(\theta_j)$.
- $BLP(\mu(\theta_j)) = z_j\bar{X}_j + (1-z_j)m$, gdzie $\bar{X}_j = \sum_{i=1}^{n_j} \frac{w_{ji}}{w_{j.}} X_{ji}$, $z_j = \frac{w_{j.}a^2}{w_{j.}a^2+s^2}$ - współczynnik zaufania
  $BLUE(m) = \sum_j \frac{z_j}{z.}\bar{X}_j$ (trzeba znać s^2 i a^2)
  $BLUP(\mu(\theta_j)) = z_j\bar{X}_j + (1-z_j\hat{m}$ (jw.)
  $EBLUP(\mu(\theta_j)) = \hat{z}_j\bar{X}_j + (1-\hat{z}_j)\tilde{m}$, gdzie $z_j = \frac{w_{j.}\hat{a}^2}{w_{j.}\hat{a}^2+\hat{s}^2}$, $\tilde{m} = \sum_j \frac{\hat{z}_j}{\hat{z}.}\bar{X}_j$
  ($\tilde{m}$ - m z dwoma daszkami).

81. **Estymacja komponentów wariancyjnych**

- $SSW = \sum_j\sum_i w_{ji}(X_{ji}-\bar{X}_j)^2$, $E(SSW) = (n.-p)s^2$, nieobciążony estymator s^2: $\hat{s}^2 = \frac{1}{n.-p}SSW$
- $SSB = \sum_j w_{j.}(\bar{X}_j - \bar{X})^2$, gdzie $\bar{X} = \sum_j \frac{w_{j.}}{w_{..}}\bar{X}_j = \sum_j\sum_i \frac{w_{ji}}{w_{..}}X_{ji}$, $\hat{a}^2 = \frac{w_{w..}}{w_{..}-\sum_j w_j^2}(SSB - \frac{p-1}{n.-p}SSW)$
- metoda oparta na pseudoestymatorach (REML) $\tilde{a}^2 = \frac{1}{p-1}\sum_{j=1}^{p} z_j(\bar{X}_j - \hat{m})^2$, $E\tilde{a}^2 = a^2$.

82. **Modele liniowe mieszane**:
$X_{ji} = m + (\mu(\theta_j) - m) + (X_{ji} - \mu(\theta_j)) = m + u_j + \varepsilon_{ji}$, gdzie:
$Eu_j = 0$, $E\varepsilon_{ji} = 0$, u z ε nieskorelowane i między sobą też, $Varu_j = a^2$, $Var\varepsilon_{ji} = \frac{s^2}{w_{j.}}$.
Model rozkłada się na addytywne efekty: m (odpowiada za średnią ogólną - liczba); u_j (odpowiada za odchyłki wierszowe - zmienna losowa); ε_{ji} (odpowiada za odchyłki pojedynczej obserwacji z wiersza - błąd losowy)

83. **Model jednokierunkowej klasyfikacji z efektami losowymi, mieszany model liniowy** $y = X\beta + Zu + \varepsilon$.
Po skalaryzacji: $l^T\beta + k^Tu = \sum_{i=1}^{p} l_i\beta_i + \sum_{j=1}^{q} k_ju_j = \mu$
Szukamy $BLP(\mu)$, $BLUP(\mu)$, $BLUE(l^T\beta)$ a potem $BLP(\mu)$ (lemat z ćwiczeń).

Wykład 12, *03.01.2011*

84. **Bayesowska rekonstrukcja obrazów** (image restoration) - (S, E), gdzie S - zbiór wierzchołków (piksli), E - zbiór krawędzi, $s \in S$, $t \in S$, $s \sim t \equiv \{s, t\} \in E$, $\delta t = \{s \sim t\}$.
A - alfabet kolorów, $x : S \to A$, x - konfiguracja (obraz), $x = (x_s, s \in S)$, $x \in X = A^s$.

85. **Rozkład Gibbsa** $\pi(x) = \frac{1}{Z_\beta} e^{-\beta H(x)}$, $H(x)$ - energia, $\beta = 1/T$, T - temperatura, $Z_\beta = \sum_{x \in X} e^{-\beta H(x)}$.

86. **Idea**: Wybieramy π - rozkład Gibbsa, jako rozkład a priori na X. Obraz „prawdziwy", „idealny": $x \in X$; „zaszumiony", „zniekształcony": $y \in Y = B^S$.

87. **Def. Rozkład Gibbsa** z interakcjami między najbliższymi sąsiadami: $H(x) = \sum_{s \sim t} J_{st}(x_s, x_t) + \sum_s h_s(x_s)$.
$\pi(x|y) \alpha f(y|x) \pi(x)$ - rozkład a posteriori.

88. **Próbnik Gibbsa** $H(x) = \sum_{s \sim t} J(x_s, x_t) + \sum_s h_s(x_s)$, $\pi(x_s|x_{-s}) = \pi(x_s|x_{\delta_s})$.

Wykład 13, *10.01.2011*

89. **Bayesowski model rekonstrukcji obrazów:**

    - **wiarogodność:** $f(y|x) = \pi_{s\in S} f(y_s|x_s)$
    - **rozkład a priori** (rozkład Gibbsa): $\pi(x) = \frac{1}{Z_\beta}\exp[-\beta\sum_{s\sim t} J(x_s, x_t)]$
    - **rozkład a posteriori**: $\pi(X|y) = \frac{1}{Z_\beta^p} exp[-\beta\sum_{s\sim t} J(x_s, x_t) + \sum_s h(x_s|y_s)]$ $(h(x_s|y_s) = h_s(x_s))$, $\pi(x|y) = \pi(x)\cdot f(y|x)$, $h(x_s|y_s) = log f(y_s|x_s)$ (było:$\sum_s h_s(x_s) =: \sum_s ln f(y_s|x_s) = log f(y|x))$.

90. **Cel - obliczanie rozkładu a posteriori:** $\pi(x_s = a|x_{-s}, y) = \pi(x_s = a|x_{-s}) = \frac{1}{Z_{\beta(s)}} exp[-\beta\sum_{t:t\sim s} J(a, x_t) + h_s(a)] = \pi(x_s = a|x_{\delta s})$.

91. **Próbnik Gibbsa** - podstawowe narzędzie w statystyce bayesowskiej. Należy do rodziny MCMC (Markov Chain Monte Carlo).

92. **MCMC - podstawy teoretyczne**
    X - dowolny (skończony dla uproszczenia), π - rozkłąd „docelowy" (a posteriori, $\pi(x) = P(X = x)$, $x \in X$)

    - **Generujemy łańcuch Markowa** $X_0, X_1, \ldots, X_n, \ldots$ na X taki że $P(X_n = x) \to_{n\to\infty} \pi(x)$, $X_n \to_{n\to\infty} \pi$.
    - **Własność Markowa** $P(X_{n+1} = y|X_n = x, X_{n-1}, \ldots, X_0 = x_0) = P(X_{n+1} = y|X_n = x) = P(x, y)$. W klasycznym MCMC posługujemy się jednorodnymi łańcuchami Markowa.
      $P(x, y)$ - prawdopodobieństwo przejścia,
      X - przestrzeń konfiguracji.
    - **Def.** π jest **rozkładem stacjonarnym** dla P, jeśli $\pi^T P = \pi^T$, tzn. $\pi(y) = \sum_{x\in X}\pi(x)P(x, y)$ $(X_n \sim \pi \Rightarrow \pi$, $P(X_{n+1} = y) = \sum_{x\in X} P(X_n = x)\cdot P(X_{n+1} = y|X_n = x))$.
    - **Uwaga** Jeśli $X_0 \sim \pi$ to $X_0, X_1, \ldots, X_n, \ldots$ jest **procesem stacjonarnym**, w szczególności $X_n \sim \pi$.
    - **Twierdzenie (ergodyczne)** Jeżeli X jest skończona, π jest rozkłądem stacjonarnym, łańcuch jest nieprzywiedlny i nieokresowy to $P(X_n = x) \to_{n\to\infty} \pi(x)$ $\forall x$ dla dowolnego rozkładu początkowego $X_0 \sim \pi_0$.
    - **Twierdzenie odwrotne** Jeśli $P(X_n = x) \to_{n\to\infty} \pi_\infty(x)$ $\forall x$, to $\pi^T P = \pi_\infty^T$ (rozkład stacjonarny).
    - P jest **nieprzywiedlny**, jeśli $\forall x, y \in X$ $\exists n$ $P^n(x, y) > 0$ $(P^n(x, y) = P(X_n = y|X_0 = x))$.
    - Łańcuch nieprzywiedlny jest **nieokresowy**, jeśli $\forall x, y \in X$ $\exists n_0 \forall n \geqslant n_0$ $P^n(x, y) > 0$ (można dojść po dowolnej liczbie kroków, byle dostatecznie dużej).

- **Twierdzenie MPWL** $g : X \to \mathbb{R}$
  $\frac{1}{n}\sum_{i=0}^{n-1} g(X_i) \to_{n\to\infty} E_\pi g = \sum_{x\in X} g(x)\pi(x)$ p.n.
  dla dowolnego rozkładu początkowego.
- **Twierdzenie CTG** $\sqrt{n}(\frac{1}{n}\sum_{i=0}^{n-1} g(x_i) - E_\pi g) \to_d N(0, \sigma^2_{as}(g))$.
  $\sigma^2_{as}(g) = Var_\pi g(X) + 2\sum_{n=1}^{\infty} Cov_\pi(g(X_0, g(X_n))$.
- **Def.** P jest **π-odwracalna**, jeśli $\pi(x)P(x,y) = \pi(y)P(y,x)$.
- Łańcuch stacjonarny jest **odwracalny**, jeśli:
  $P(X_n = x, X_{n+1} = y) = P(X_n = y, X_{n+1} = x)$
  $(X_0, X_1, \ldots, X_n) =^d (X_n, X_{n-1}, \ldots, X_0)$.
- **Stwierdzenie** Jeśli P jest π-odwracalna to $\pi^T P = \pi^T$.

93. **Algorytm Metropolisa-Hastingsa**

- Mamy prawdopodobieństwa przejścia $Q = (Q(x,y))$ takie, że umiemy generować łańcuch Markowa o przejściach Q. Zakładamy symetrię $Q(x,y) = Q(y,x)$. Rozkład $U(X)$ (jednostajny) jest stacjonarny.
- **Krok algorytmu** ($X_n = x$):
  1) Gen $Y \sim Q(x,\cdot)$ - generowanie „propozycji”
  2) Obliczamy $a(x,Y) = \frac{\pi(Y)}{\pi(x)} \wedge 1$
  3) z prawdopodbieństwem $a(x,Y)$ bierzemy $X_{n+1} := Y$ (akceptacja), $1 - a(x,Y)$ bierzemy $X_{n+1} := x$ (odrzucenie).
- **Modyfikacja Hastingsa:** Q nie musi być symetryczna:
  $a(x,Y) = \frac{\pi(Y)Q(Y,X)}{\pi(X)Q(x,Y)} \wedge 1$.
- Łańcuch Metropolisa-Hastingsa ma prawdopodobieństwa przejścia
  $P(x,y) = Q(x,y)a(x,y)$ $(y \neq x)$.
- **Twierdzenie** P jest π-odwracalny.
- **Wniosek** Jeśli Q jest nieprzywiedlna i π nie jest jednostajny to łańcuch Metropolisa-Hastingsa zmierza do π: $P(X_n = x) \to_{n\to\infty} \pi(x)$.
- **Uwaga** Jeśli $P(x,x) > 0$ (po jednym kroku) to łańcuch jest nieokresowy.
- **Uwaga** Jeżeli $\pi_\beta(x) = \frac{1}{Z_\beta} e^{-\beta H(x)}$ i $\beta \to \infty$, to
  $\pi_\beta(x) = \frac{1}{|X_{min}|}$, gdy $H(x) = H_{min}$ lub 0 gdy $H(x) > H_{min}$.
  $H_{min} = \min_{x\in X} H(x)$, $X_{min} = \{x : H(x) = H_{min}\}$.
- **Symulowane wyżarzanie** (simulated annealing, $Met + \beta \to \infty$).

Wykład 14, *17.01.2011*

94. **Bayesowski model rekonstrukcji obrazów**:
$x \in X = A^S$, $x = (x_s)_{s \in S}$, $\pi(x) = \frac{1}{Z} e^{-H(x)}$
$H(x) = \beta \sum_{s \sim t} J(x_s, x_t) + \sum_s h(x_s|y_s)$
$h(x_s|y_s) = log f(y_s|x_s)$, $J(x_s, x_t) = \psi(|x_s - x_t|)$.

95. **Szum Piossonowski**:
- $x_s \in \mathbb{R}_+ = (0, \infty)$, $X = \mathbb{R}_+^s$, x_s – intensywności rozkładu Poissona
- $y_s|x_s \sim Poiss(x_s)$, $f(y_s|x_s) = e^{-x_s} \frac{x_s^{y_s}}{y_s!}$, $y_s \in \{0, 1, \ldots\}$, $h(x_s|y_s) = -x_s + y_s log x_s$
- $x_s = \lambda(x_s)$, $\lambda = (\lambda(1), \ldots, \lambda(k))$ – paleta intensywności, $z_s \in \{1, \ldots, k\}$ – etykiety intensywności
- wprowadza się dodatkowe zmienne, które ułatwiają obliczenia; zmienne pomocnicze (z, λ), $z = \varepsilon^s$ – etykietki umieszczenie etykietek w węzłach; $\lambda = \mathbb{R}_+^\varepsilon$ – umieszczenie intensywności w etykietkach, $x_s = \lambda(z_s)$
- **energia Pottsa**=energia a priori (rozkład a priori na z)+energia a posteriori; rozkład a priori na λ jest rozkładem jednostajnym (na prostej) $H(z, \lambda) = \beta \sum_{st} 1(z_s \neq z_t) + \sum_s [-\lambda(z_s) + y_s log \lambda(z_s)]$
- **Próbnik Gibbsa** $z \sim \pi(z|\lambda)$ $(\pi(z_s|z_{-s}, \lambda))$, $\lambda \sim \pi(\lambda|z)$, λ ustalone.

96. **PET** - positron emission tomography.