

Statystyka Bayesowska, Egzamin II, luty 2019, MIM UW

1. Pewien test laboratoryjny daje zawsze wynik dodatni jeśli badana osoba jest zarażona wirusem „W”. Jeśli badana osoba nie jest zarażona wirusem „W” to test daje omyłkowo wynik dodatni z prawdopodobieństwem $2/9$, a z prawdopodobieństwem $7/9$ daje wynik ujemny. Wiadomo, że w populacji osób zgłaszających się na badanie, $1/10$ jest zarażonych wirusem „W”.
 - (a) Oblicz prawdopodobieństwo, że dla losowo wybranej osoby test da wynik dodatni.
 - (b) Oblicz prawdopodobieństwo, że osoba jest zarażona wirusem „W”, jeśli wiadomo, że test dał wynik dodatni.
 - (c) Oblicz prawdopodobieństwo, że powtórne badanie, przeprowadzone *niezależnie dla tej samej osoby* da wynik dodatni, jeśli wiadomo, że pierwsze badanie dało wynik dodatni.
2. Zmienne losowe $X_1, \dots, X_n$ są, przy danym θ , warunkowo niezależne i każda z nich ma rozkład jednostajny $U(0, \theta)$. Parametr θ jest zmienną losową o rozkładzie *a priori* jednostajnym $U(0, 1)$.
 - (a) Wyznacz rozkład *a posteriori* $\pi(\theta|x_1, \dots, x_n)$.
 - (b) Oblicz estymator Bayesowski parametru θ przy kwadratowej funkcji straty, $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$.
 - (c) Oblicz estymator Bayesowski parametru θ przy „względnej kwadratowej” funkcji straty, $L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2/\theta^2$.
3. Niech Z_1 i Z_2 będą niezależnymi zmiennymi losowymi o normalnym rozkładzie prawdopodobieństwa $N(0, 1)$. Niech $X = 2 + Z_1 - Z_2$ i $Y = -1 + Z_1 + 2Z_2$.
 - (a) Oblicz BLP($Y|X$), czyli podaj współczynniki β_0^* i β^* takie aby zmienna losowa $\beta_0^* + \beta^*X$ miała najmniejszy błąd średniokwadratowy predykcji:

$$\mathbb{E}(Y - (\beta_0^* + \beta^*X))^2 = \min_{\beta_0, \beta} \mathbb{E}(Y - (\beta_0 + \beta X))^2.$$

- (b) Oblicz $\mathbb{E}(Y|X)$.
- (c) Oblicz $\text{Var}(Y|X)$.

4. Rozważmy następujący hierarchiczny model Bayesowski.

- $X \sim \text{Bin}(n, \theta)$, czyli $p(x|\theta) = \mathbb{P}_\theta(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$ dla $x = 0, 1, \dots, n$.
- $\theta \sim \text{Beta}(\psi, 1)$, czyli $p(\theta|\psi) = \psi \theta^{\psi-1}$ dla $0 < \theta < 1$.
- $\psi \sim \text{Gamma}(\alpha, \lambda)$, czyli $p(\psi) = \lambda^\alpha \Gamma(\alpha)^{-1} \psi^{\alpha-1} e^{-\lambda\psi}$ dla $\psi > 0$.

Zakładamy warunkową niezależność: $p(x|\theta, \psi) = p(x|\theta)$.

Obserwujemy zmienną losową X . Parametry α, λ i n są znane. Zaprojektuj próbnik Gibbsa zbieżny do rozkładu *a posteriori* $p(\theta, \psi|x)$:

- (a) Podaj rozkład $p(\theta|\psi, x)$,
- (b) Podaj rozkład $p(\psi|\theta, x)$.

5. Rozważmy następujący model klasyfikacji (bez zawieszania decyzji). Obserwujemy zmienną losową X .

W klasie I, czyli dla $\theta = 1$, mamy $X \sim \text{Ex}(2)$, czyli X ma rozkład wykładniczy o gęstości $f_1(x) = 2e^{-2x}$ dla $x \geq 0$;

W klasie II, czyli dla $\theta = 2$, mamy $X \sim \frac{1}{2}\text{Ex}(2) + \frac{1}{2}\text{Ex}(1)$, czyli rozkład zmiennej X jest mieszkanką dwóch rozkładów wykładniczych i ma gęstość $f_2(x) = e^{-2x} + \frac{1}{2}e^{-x}$ dla $x \geq 0$.

Prawdopodobieństwo *a priori* tego, że X pochodzi z klasy I jest równe $\mathbb{P}(\theta = 1) = 1/2$.

- (a) Podaj klasyfikator Bayesowski, czyli funkcję decyzyjną $\delta^* : \mathbb{R} \rightarrow \{1, 2\}$ minimalizującą ryzyko dla „0-1” funkcji straty $L(\theta, a)$ określonej następująco: $L(\theta, a) = \mathbb{I}(\theta \neq a)$.
- (b) Oblicz ryzyko Bayesowskie klasyfikatora δ^* , czyli prawdopodobieństwo błędnej klasyfikacji $\mathbb{P}(\delta^*(X) \neq \theta)$.