Podejmowanie decyzji statystycznych

12.1. Sprzedawca jest zainteresowany maksymalizacją zysku ze sprzedaży pewnego, łatwo psującego się towaru. Zakup jednego opakowania kosztuje go \$20, zaś sprzedaje je za \$50. Po jednym dniu niesprzedany towar należy wyrzucić jako niezdatny do spożycia. W ciągu stu dni sprzedawca zaobserwował, że klienci kupowali następujące ilości towaru:

Ilość towaru	10	11	12	13
Ilość dni	15	20	40	25

Ile towaru powinien sprzedawca zamówić, by mógł oczekiwać największego zysku?

12.2. Przedsiębiorca zainteresowany jest zatrudnieniem w swojej firmie samochodowej kilku mechaników. Na kolejny rok przewidywane są następujące ilości godzin pracy dla mechaników:

Ilość godzin	10000	12000	14000	16000
Prawdopodobieństwo	.2	.3	.4	.1

Planowana zapłata za jedną godzinę pracy mechanika wynosi \$9 zaś spodziewany zysk z jednej godziny pracy — \$16. Mechanik może pracować 40 godzin w tygodniu oraz ma prawo do dwutygodniowego urlopu. Na podstawie podanych informacji podać optymalną liczbę mechaników, która powinna być zatrudniona.

12.3. Przedsiębiorstwo lotnicze otrzymuje propozycję zakupienia dziesięciu używanych samolotów, przy czym wszystkie są mniej więcej w takim samym stanie technicznym. Pewna nieznana liczba θ tych samolotów może latać 1000 godzin bez potrzeby większych napraw i każdy z takich samolotów przyniesie zysk w wysokości $1000p$; każdy z pozostałych samolotów dozna jakiegoś poważniejszego uszkodzenia w ciągu pierwszych 1000 godzin eksploatacji, co przyniesie przedsiębiorstwu stratę w wysokości $1000q$. Należy zdecydować, czy przyjąć czy odrzucić ofertę. Przed podjęciem decyzji przedsiębiorstwo może otrzymać pewne dalsze informacje, a mianowicie za cenę $1000r$ może otrzymać jeden samolot do prób, eksploatować go przez 1000 godzin i uzależnić swoją decyzję od tego, czy ten samolot latał 1000 godzin bez awarii, czy nie. Podać wszystkie reguły decyzyjne i ich ryzyka. Wybrać optymalną regułę decyzyjną.

12.4. Pewna osoba zamierza sprzedawać napój *Migotka* w trakcie meczu piłki nożnej i musi zawczasu zdecydować o wielkości zamówienia. Przypuśćmy, że na każdym sprzedawanym w trakcie gry litrze zyskuje m, zaś traci c na każdym zamówionym, lecz nie sprzedanym. Załóżmy, że popyt na *Migotkę* w trakcie gry, mierzony w litrach, jest ciągłą zmienną losową X o funkcji gęstości f i dystrybuancie F. Przy jakiej wysokości zamówienia oczekiwany zysk będzie maksymalny?

12.5. Rozważmy zagadnienie decyzyjne, w którym $\Theta = \{\theta_1, \theta_2\}$ oraz $\mathcal{D} = \{d_1, d_2, d_3\}$, funkcja straty określona jest w następujący sposób:

	d_1	d_2	d_3
θ_1	0	10	3
θ_2	10	0	3

Niech obserwowana zmienna losowa X ma rozkład

$$P(X = 1|\theta = \theta_1\} = 0.75;\; P(X = 0|\theta = \theta_1\} = 0.25; \qquad P(X = 1|\theta = \theta_2\} = 0.25;\; P(X = 0|\theta = \theta_2\} = 0.75$$

Zmienna losowa może być obserwowana n krotnie. Niech $P\{\theta = \theta_1\} = p = 1-P\{\theta = \theta_2\}$ $(0 \leq p \leq 1)$. Określić bayesowską regułę decyzyjną na podstawie obserwacji $X_1, \ldots, X_n$ i naszkicować jej ryzyko jako funkcję prawdopodobieństwa p. Zakładając, że koszt każdej obserwacji wynosi c wyznaczyć optymalną wielkość n. Wyznaczyć optymalną wielkość próby, gdy koszt obserwacji o wartości 1 wynosi c_1, zaś koszt obserwacji o wartości 0 wynosi 0.

12.6. Niech $X_1, X_2, \ldots, X_n$ będzie próbką losową z rozkładu $N(\mu, \sigma^2)$ o nieznanych μ i σ^2 i niech rozkład *a priori* tych parametrów będzie taki, że μ i $\log\sigma$ są niezależne, a brzegowy rozkład każdej z tych wielkości jest jednostajny. Pokazać, że estymatorem bayesowskim parametru μ przy kwadratowej funkcji strat $L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2$ jest $\bar{X}$. Ile wynosi ryzyko bayesowskie tego estymatora?

12.7. Niech X będzie zmienną losową o rozkładzie $N(\theta, 1)$ i niech strata spowodowana szacowaniem θ za pomocą $\hat{\theta}(X)$ ma postać

$$\begin{aligned} &a\{\hat{\theta}(X) - \theta\}, \quad \text{gdy} \quad \hat{\theta}(X) \geq \theta, \\ &b\{\theta - \hat{\theta}(X)\}, \quad \text{gdy} \quad \hat{\theta}(X) < \theta, \end{aligned}$$

gdzie $a > 0$ oraz $b > 0$. Pokazać, że ryzyko estymatora $\hat{\theta}_k$ postaci

$$\hat{\theta}_k(X) = X - k$$

wyraża się wzorem

$$(a + b)\{\varphi(k) + k\Phi(k)\} - ka,$$

gdzie φ i Φ są odpowiednio gęstością i dystrybuantą rozkładu $N(0, 1)$. Następnie pokazać, że w klasie $\{\hat{\theta}_k : k \text{ rzeczywiste}\}$ istnieje estymator o jednostajnie minimalnym ryzyku i że to minimalne ryzyko jest równe

$$(a + b)\varphi\left[\Phi^{-1}\left(\frac{a}{a + b}\right)\right].$$

12.8. Bochenek chleba musi ważyć co najmniej w gramów. W pewnej piekarni ciężar chleba w dużym wypieku jest zmienną losową o rozkładzie $N(\mu, 1/\tau)$. Parametr μ jest wielkością regulowaną, natomiast τ zmienia się od wypieku do wypieku według rozkładu chi–kwadrat o ν stopniach swobody. Koszt produkcji bochenka chleba, gdy średni ciężar jest równy μ wynosi $k + l\mu$, a cena zbytu bochenka o prawidłowym ciężarze jest równa m. Bochenki o zbyt małym ciężarze nie przynoszą zysku. Pokazać, że średni zysk na bochenku, gdy wypiek ustawiony jest na wielkość μ, wynosi

$$m\Psi\{(\mu - w)\sqrt{n}\} - (k + l\mu),$$

gdzie Ψ jest rozkładem t o ν stopniach swobody. Wyznaczyć na tej podstawie najlepszą wartość μ.

12.9. W procesie mierzenia zawartości RNA w pewnych komórkach pojawia się trudność związana z tym, że dwie komórki mogą znajdować się tak blisko siebie, iż stają się nierozróżnialne. Wykonuje się dwa niezależne pomiary X_1 i X_2, przy czym każdy z nich może odnosić się do jednej lub dwóch (niezależnych) komórek. Należy zdecydować, czy dana para pomiarów dotyczy dwóch pojedynczych komórek, jednej pojedynczej komórki i jednej pary, czy też dwóch par komórek.
Przypuśćmy, iż wiadomo, że zawartość pojedynczej komórki, mierzona w odpowiednich jednostkach, ma rozkład o gęstości xe^{-x} $(x \geq 0)$ i że prawdopodobieństwo *a priori* tego, że pomiar dotyczy dwóch komórek zamiast jednej wynosi π. Przypuśćmy również, że strata jest równa zeru, gdy decyzja jest prawidłowa i jest równa jedności, gdy decyzja jest błędna. Naszkicować na płaszczyźnie (x_1, x_2) rozwiązanie bayesowskie tego zagadnienia.

12.10. Za pomocą pewnego odbiornika rejestruje się n kolejnych pomiarów $X_1, X_2, \ldots, X_n$ intensywności sygnału radiowego. Jeżeli był nadawany pewien sygnał, to wynik pomiaru ma postać $X_j = a_j + \varepsilon_j$, gdzie $a_1, a_2, \ldots, a_n$ są znane, a losowe zakłócenia $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ są realizacją wielowymiarowej zmiennej losowej o rozkładzie normalnym ze średnią równą zeru i z macierzą kowariancji V. Jeżeli sygnał nie był nadawany, to $X_j = \varepsilon_j$ $(j = 1, 2, \ldots, n)$. Należy zdecydować, czy sygnał został rzeczywiście nadany. Prawdopodobieństwo *a priori* tego zdarzenia wynosi p. Straty związane z nieprawidłowym orzeczeniem, że sygnał został lub nie został nadany, są równe odpowiednio $L - 1$ i L_2. Wyznaczyć regułę decyzyjną realizującą minimum oczekiwanych strat.
Przypuśćmy, że decyzję podejmuje się zgodnie z tą optymalną regułą i że $pL_2 = (1 - p)L_1$. Jak wygląda optymalny ciąg $\{a_i\}$ sygnałów, jeżeli analiza mocy transmisji prowadzi do ograniczenia $\sum a_i^2 = 1$?

12.11. W celu podjęcia decyzji, którą z dwóch odmian pszenicy wprowadzić do masowej produkcji, wykonuje się eksperyment polegający na tym, że każdą z odmian bada się na n poletkach doświadczalnych. Obserwowane plony X_{ij} $(i = 1, 2, j = 1, 2, \ldots, n)$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym ze średnimi μ_i i wariancją σ^2, przy czym rozkład *a priori* parametru (μ_1, μ_2) jest taki, że μ_1 i μ_2 są niezależne i mają rozkłady brzegowe normalne o średnich równych odpowiednio α_1 i α_2 i wariancji σ_0^2. Wyznaczyć rozkład *a posteriori* (μ_1, μ_2) przy założeniu, że wariancja σ^2 jest znana. Pokazać, że gdy funkcje strat mają postać $L_i = -k\mu_i$ $(i = 1, 2)$, wtedy oczekiwane ryzyko osiąga minimum dla decyzji: wybrać odmianę 1, gdy $X_{1\cdot} - X_{2\cdot} > c$ lub odmianę 2, gdy $X_{1\cdot} - X_{2\cdot} < c$, gdzie $nX_{i\cdot} = \sum_j X_{ij}$ oraz $c = (\alpha_2 - \alpha_1)\sigma^2/(n\sigma_0^2)$.

12.12. Niech (X_n) będzie ciągiem niezależnych zmiennych losowych o rozkładzie Poissona z jednakową średnią θ. W celu oszacowania θ obserwuje się ten ciąg sekwencyjnie. Zakładamy, że: początkowa wiedza o θ opisana jest za pomocą niewłaściwego rozkładu *a priori* o stałej gęstości na przedziale $(0, \infty)$; strata spowodowana oszacowaniem θ przez $\hat{\theta}$ wynosi $(\hat{\theta} - \theta)^2$; koszt każdej obserwacji jest równy c. Udowodnić, że optymalna bayesowska sekwencyjna reguła decyzyjna ma postać: przerwać obserwacje, gdy tylko $\sum_{i=1}^n X_i + 1 < cn^2(n+1)$ i za oszacowanie θ przyjąć $\sum_{i=1}^n (X_i + 1)/n$.

12.13. Niech $X_1, X_2, \ldots, X_{10}$ będzie próbką losową z rozkładu normalnego $N(0, \theta)$. Wyznaczyć rozkład *a posteriori* parametru θ, gdy rozkład *a priori* tego parametru jest taki sam jak rozkład zmiennej losowej $\theta_0 \nu_0 / \chi^2(\nu_0)$, gdzie θ_0 i ν_0 są dodatnimi stałymi.
Wyznaczyć najlepszy bayesowski estymator punktowy $\hat{\theta}$, gdy funkcja straty ma postać $(\hat{\theta} - \theta)^2$. Wykazać, że gdy $\nu_0 \to 0$ przy stałym θ_0, wtedy

$$\hat{\theta} \to \frac{1}{8} \sum X_i^2.$$

Porównując estymator $\frac{1}{8} \sum X_i^2$ z estymatorem różniącym się od niego pewnym mnożnikiem pokazać, że jest to estymator niedopuszczalny.