Aufgabe 4

Betrachten Sie das Poisson-Modell, d.h. $X \sim \text{Po}(\lambda)$ und für den Parameter λ wird eine $\text{Ga}(a,b)$-Priori-Verteilung angenommen.

(a) Berechnen sie die Posteriori-Verteilung $p(\lambda|X)$ explizit, d.h. inklusive Normierungskonstante.

Hinweis: $\Gamma(x) = \int\limits_0^\infty t^{x-1}\exp(-t)\,dt$.

(b) Warum genügt es, die Posteriori nur bis auf eine multiplikative Konstante zu bestimmen?

Lösung

a) Es gilt:

$$\begin{aligned} X \sim \text{Po}(\lambda) \quad &\Rightarrow \quad p_X(x|\lambda) = \frac{\lambda^x}{x!}\exp(-\lambda) \\ \lambda \sim \text{Ga}(a,b) \quad &\Rightarrow \quad p_\lambda(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}\exp(-b\lambda) \end{aligned}$$

Bestimmung der Posteriori

$$\begin{aligned} p(\lambda|x) &= \frac{p_X(x|\lambda)\,p_\lambda(\lambda)}{\int_0^\infty p_X(x|y)\,p_\lambda(y)\,dty} \\ &= \frac{\frac{\lambda^x}{x!}\exp(-\lambda)\frac{b^a}{\Gamma(a)}\lambda^{a-1}\exp(-b\lambda)}{\int_0^\infty \frac{y^x}{x!}\exp(-y)\frac{b^a}{\Gamma(a)}y^{a-1}\exp(-by)\,dty} \\ &= \frac{\lambda^{x+a-1}\exp(-\lambda[b+1])}{\int_0^\infty y^{x+a-1}\exp(-y[b+1])\,dty} \\ &= \frac{\lambda^{x+a-1}\exp(-\lambda[b+1])}{\int_0^\infty \left(\frac{y(b+1)}{(b+1)}\right)^{x+a-1}\exp(-\underbrace{y[b+1]}_{=:z})\,dty} \\ &= \frac{\lambda^{x+a-1}\exp(-\lambda[b+1])}{\left(\frac{1}{(b+1)}\right)^{x+a-1}\int_0^\infty z^{x+a-1}\exp(-z)\frac{1}{b+1}\,dtz} \\ &= \frac{\lambda^{x+a-1}\exp(-\lambda[b+1])}{\left(\frac{1}{(b+1)}\right)^{x+a}\int_0^\infty z^{x+a-1}\exp(-z)\,dtz} \\ &= \frac{\lambda^{x+a-1}\exp(-\lambda[b+1])}{\left(\frac{1}{(b+1)}\right)^{x+a}\Gamma(x+a)} \\ &= \frac{(b+1)^{x+a}}{\Gamma(x+a)}\lambda^{x+a-1}\exp(-\lambda[b+1]) \end{aligned}$$

Dies ist die Dichte von $\text{Ga}(a+x, b+1)$, d.h. $\lambda|x \sim \text{Ga}(a+x, b+1)$.

b) Betrachte allgemein $X \sim F(\theta)$ mit Dichte $p(x|\theta)$ (Likelihood). Für den Parameter θ wird ebenfalls eine Dichte angenommen: $p(\theta)$, $\theta \in \Theta$ (Priori).

Sei die Posteriori(-Dichte) nur bis auf Proportionalität bestimmt, d.h.

$$\exists\, g(\cdot|x) : \Omega \to [0,\infty) \text{ und } c > 0: \qquad p(\theta|x) = c \cdot g(\theta|x) \quad \forall\, \theta \in \Theta.$$

Notation: $p(\theta|x) \propto g(\theta|x)$.\ Die Posteriori ist eine Dichte, daher gilt

$$\begin{aligned} 1 &= \int_\Theta p(\theta|x)\ d\theta = \int_\Theta c \cdot g(\theta|x)\ d\theta = c \cdot \int_\Theta g(\theta|x)\ d\theta \\ \Leftrightarrow\ c &= \frac{1}{\int_\Theta g(\theta|x)\ d\theta} \end{aligned}$$

also ist c eindeutig durch g bestimmt.

Nach Definition der Posteriori gilt

$$p(\theta|x) = \frac{p(x|\theta)\,p(\theta)}{\int_\Theta p\left(x\middle|\tilde{\theta}\right)p(\tilde{\theta})\,dt\tilde{\theta}},$$

d.h. $g(\theta|x) = p(x|\theta)\,p(\theta)$ ("Likelihood $\cdot$ Priori") und $c = \frac{1}{\int_\Theta p(x|\tilde{\theta})p(\tilde{\theta})\,dt\tilde{\theta}}$ ist eindeutig bestimmt, falls g bekannt ist.

Aufgabe 1
Eine Schokoladenfabrik stellt Pralinen her, die jeweils eine Kirsche enthalten. Die benötigten Kirschen werden an zwei Maschinen entkernt. Maschine A liefert 70 % dieser Kirschen, wobei 8 % der von A gelieferten Kirschen den Kern noch enthalten. Maschine B produziert 30 % der benötigten Kirschen, wobei 5 % der von B gelieferten Kirschen den Kern noch enthalten. Bei einer abschließenden Gewichtskontrolle werden 95 % der Pralinen, in denen ein Kirschkern enthalten ist, aussortiert, aber auch 2 % der Pralinen ohne Kern.

(a) Modellieren Sie diesen mehrstufigen Vorgang geeignet. Wie groß ist die Wahrscheinlichkeit, dass eine Praline mit Kirschkern in den Verkauf gelangt?

(b) Ein Kunde kauft eine Packung mit 100 Pralinen. Wie groß ist die Wahrscheinlichkeit, dass nur gute Pralinen, also Pralinen ohne Kirschkern, in der Packung sind?

Lösung

a) Wir modellieren die Situation durch ein dreistufiges Experiment mit dem Grundraum $\Omega := \Omega_1 \times \Omega_2 \times \Omega_3$ sowie der Festsetzung

$$\mathbb{P}\left(\{(\omega_1, \omega_2, \omega_3)\}\right) := p_1(\omega_1) \cdot p_2(\omega_1, \omega_2) \cdot p_3(\omega_1, \omega_2, \omega_3)$$

für $(\omega_1, \omega_2, \omega_3) \in \Omega$. Dabei werden die drei Stufen wie folgt beschrieben:

- $\Omega_1 := \{A, B\}$ (Kirsche von Maschine A bzw. B) und $p_1(A) := 0.7, p_1(B) := 0.3$.
- $\Omega_2 := \{mK, oK\}$ (Kirsche mit Kern, Kirsche ohne Kern) und

$$p_2(A, mK) := 0.08, \quad p_2(A, oK) := 0.92,$$
$$p_2(B, mK) := 0.05, \quad p_2(B, oK) := 0.95.$$

- $\Omega_3 := \{V, nV\}$ (Praline im Verkauf, Praline nicht im Verkauf) und

$$p_3(\omega_1, mK, V) := 0.05, p_3(\omega_1, mK, nV) := 0.95,$$
$$p_3(\omega_1, oK, V) := 0.98, p_3(\omega_1, oK, nV) := 0.02.$$

Man beachte, dass p_3 nicht von ω_1 abhängt. Sei $C := \{(A, mK, V), (B, mK, V)\}$ das Ereignis, dass eine Praline mit Kirschkern in den Verkauf gelangt. Dann ist

$$\mathbb{P}(C) = 0.7 \cdot 0.08 \cdot 0.05 + 0.3 \cdot 0.05 \cdot 0.05 = 0.00355.$$

Mit 0.355-prozentiger Wahrscheinlichkeit gelangt also eine Praline mit Kirschkern in den Verkauf.
b) Wir müssen zunächst berechnen, mit welcher Wahrscheinlichkeit eine Praline ohne Kirschkern in den Verkauf gelangt, d. h. die Wahrscheinlichkeit des Ereignisses

$$D := \{(A, oK, V), (B, oK, V)\}.$$

Diese ist gegeben durch

$$\mathbb{P}(D) = 0.7 \cdot 0.92 \cdot 0.98 + 0.03 \cdot 0.95 \cdot 0.98 = 0.91042.$$

Folglich ist die Wahrscheinlichkeit, dass eine in den Verkauf gelangte Praline keinen Kirschkern enthält, gleich

$$\begin{aligned}\mathbb{P}(oK \mid V) &= \frac{\mathbb{P}(oK \cap V)}{\mathbb{P}(V)} = \frac{\mathbb{P}(oK \cap V)}{\mathbb{P}(mK \cap V) + \mathbb{P}(oK \cap V)} \\ &= \frac{\mathbb{P}(D)}{\mathbb{P}(C) + \mathbb{P}(D)} = \frac{91042}{91397} \approx 0.9961.\end{aligned}$$

ergibt sich die Wahrscheinlichkeit p, dass in einer Packung mit 100 Pralinen nur gute

$$p = \left(\frac{91042}{91397}\right)^{100} \approx 0.6776.$$

Aufgabe 2

Nehmen wir an, die Prävalenz von Corona an einem gewissen Tag liege bei 20 ansteckenden Personen bei 100.000 Einwohnern.

Die Spezifität (Wahrscheinlichkeit, dass der Test einer gesunden Person negativ ausfällt) eines Anti-Gen-Tests liege bei 98%. Die Sensitivität (Wahrscheinlichkeit, dass der Test bei einer erkrankten Person positiv ausfällt) des Anti-Gen-Tests liege bei 90%.

Die Spezifität eines PCR-Tests liege bei 99%. Die Sensitivität des PCR-Tests liege bei 98%.

Es sei folgendes Vorgehen üblich: Es wird erst ein Anti-Gen-Test durchgeführt. Fällt er positiv aus, so wird ein PCR-Test ausgeführt.

(a) Formulieren Sie die hier dargestellten Ereignisse und Wahrscheinlichkeiten.

Lösung

- C: Person ist Corona-infiziert.
- A: Anti-Gen-Test ist positiv.
- P: PCR-Test ist positiv.

Angaben:

- Sensitivität Anti-Gen-Test: $P(A|C) = 0.9$
- Spezifität Anti-Gen-Test: $P(\bar{A}|\bar{C}) = 0.98$
- Sensitivität PCR-Test: $P(P|C) = 0.98$
- Spezifität PCR-Test: $P(\bar{P}|\bar{C}) = 0.99$

Priori-Annahme: $P(C) = 20/100000 = 0.0002 = 0.02\%$

(b) Wie hoch ist die Wahrscheinlichkeit, dass eine Person, bei der auch der PCR-Test positiv ist, tatsächlich Corona-infiziert ist?

Lösung

Zuerst-Anti-Gen-Test, gesucht: $P(C|A)$. Mit Formal von Bayes:

$$\begin{aligned}
P(C|A) &= \frac{P(A|C)P(C)}{P(A)} && (1)\\
&= \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|\bar{C})P(\bar{C})} && (2)\\
&= \frac{0.9 \cdot 0.0002}{0.9 \cdot 0.0002 + (1-0.98)\cdot(1-0.0002)} && (3)\\
&= \frac{0.00018}{0.020176} = 0.008921491 \approx 0,89\% && (4)\\
& && (5)
\end{aligned}$$

PCR-Test nur, wenn Anti-Gen-Test positiv. Dann ist die Priori-Information also $P(C|A) = 0.008921491$. Unter der Annahme, dass Anti-Gen-Test und PCR-Test gegeben C unabhängig sind (also $P(P|C, A) = P(P|C)$), gilt mit Bayes-Formel:

$$\begin{aligned}
P(C|A,P) &= \frac{P(P|C,A)P(C|A)}{P(P|A)} && (6)\\
&= \frac{P(P|C)P(C|A)}{P(P|C)P(C|A)+P(P|\bar{C})P(\bar{C}|A)} && (7)\\
&= \frac{0.98\cdot 0.008921491}{0.99\cdot 0.008921491+(1-0.99)\cdot(1-0.008921491)} && (8)\\
&\approx \frac{0.008743061}{0.01874306} \approx 0.4664692 \approx 46,7\% && (9)\\
& && (10)
\end{aligned}$$

(c) Wie hoch ist die Wahrscheinlichkeit, dass eine infizierte Person nicht erkannt wird?
Lösung
Person wird nicht erkannt, wenn Anti-Gen-Test negativ, obwohl Person Coron-infiziert, also $P(\bar{A}|C)$, oder wenn für eine Corona-infizierte Person Anti-Gen-Test positiv, PCR-Test negativ, also $P(A|C)\cdot P(\bar{P}|C)$.
Zusammen:

$$P(\bar{A}|C)+P(A|C)\cdot P(\bar{P}|C) = (1-0.9)+0.9\cdot(1-0.98) = 11,8\%$$

Alternativ sei folgende Vorgehen üblich: Es werden nur Personen mit starken Symptomen getestet. Das betrifft etwa 2% der Personen. Wir können annehmen, dass unter diesen etwa 1% der Personen infiziert sei.

(d) Formulieren Sie die hier dargestellten Ereignisse und Wahrscheinlichkeiten.
Lösung
Einzige Änderung zu oben, Priori-Annahme ist nun: $P(C)=0.01$. Angabe "betrifft etwa 2

(e) Wie hoch ist die Wahrscheinlichkeit, dass eine Person bei der bei diesem Vorgehen der PCR-Test positiv ist, tatsächlich Corona-infiziert ist?
Lösung
Bayes-Formel:

$$\begin{aligned}
P(C|P) &= \frac{P(P|C)P(C)}{P(P)} && (11)\\
&= \frac{P(P|C)P(C)}{P(P|C)P(C)+P(P|\bar{C})P(\bar{C})} && (12)\\
&= \frac{0.98\cdot 0.01}{0.98\cdot 0.01+(1-0.99)\cdot(1-0.01)} && (13)\\
&= \frac{0.0098}{0.0197} \approx 0.4974619 \approx 49.7\% && (14)\\
& && (15)
\end{aligned}$$

(f) Wie hoch ist die Wahrscheinlichkeit, dass eine infizierte Person nicht erkannt wird?
Lösung
$P(C|\bar{P}) = 1-0.99 = 0.01 = 1\%$. Gilt allerdings nur unter der Annahme, dass Personen ohne starke Symptome nicht infiziert sind - ansonsten ist keine Aussage möglich!