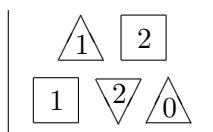


Ü b u n g s b l a t t 13

Mit * und ** gekennzeichnete Aufgaben können zum Sammeln von Bonuspunkten verwendet werden. Lösungen von *-Aufgaben sind schriftlich abzugeben im Zettelkasten Nr. 5 auf dem D1 bis Mittwoch, 4.7.07, 11:00 Uhr. Lösungen von **-Aufgaben sind per Web-Formular unter <http://www.math.upb.de/~walter> (→ Lehre SS 07 → Übungen) abzuliefern bis spätestens Mittwoch, 4.7.07, 23⁵⁹ Uhr.

Aufgabe 70: (Unabhängigkeit von Zufallsvariablen)

Aus der untenstehenden Urne wird ein Objekt gezogen. Sei X die Zahl auf dem Objekt, sei Y die Anzahl der Ecken. Bestimme $E(X)$, $\sigma(Y)$ und $E(XY)$. Sind X und Y unabhängig?



Musterlösung:

Die Verteilung von X bzw. Y ist gegeben durch

$$P(X = 0) = \frac{1}{5}, \quad P(X = 1) = \frac{2}{5}, \quad P(X = 2) = \frac{2}{5}, \quad P(Y = 3) = \frac{3}{5}, \quad P(Y = 4) = \frac{2}{5}.$$

Die gemeinsame Verteilung von X und Y ist gegeben durch

$$\begin{aligned} P(X = 0 \text{ und } Y = 3) &= \frac{1}{5}, \\ P(X = 1 \text{ und } Y = 3) &= \frac{1}{5}, \\ P(X = 2 \text{ und } Y = 3) &= \frac{1}{5}, \\ P(X = 0 \text{ und } Y = 4) &= 0, \\ P(X = 1 \text{ und } Y = 4) &= \frac{1}{5}, \\ P(X = 2 \text{ und } Y = 4) &= \frac{1}{5}. \end{aligned}$$

Die Verteilung der Produktvariable XY ist gegeben durch

$$\begin{aligned} P(XY = 0) &= P(X = 0 \text{ und } Y = 3) + P(X = 0 \text{ und } Y = 4) = \frac{1}{5}, \\ P(XY = 3) &= P(X = 1 \text{ und } Y = 3) = \frac{1}{5}, \\ P(XY = 4) &= P(X = 1 \text{ und } Y = 4) = \frac{1}{5}, \\ P(XY = 6) &= P(X = 2 \text{ und } Y = 3) = \frac{1}{5}, \\ P(XY = 8) &= P(X = 2 \text{ und } Y = 4) = \frac{1}{5}. \end{aligned}$$

Hiermit lässt sich alles berechnen, z.B.:

$$\begin{aligned}E(X) &= 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} + 2 \cdot \frac{2}{5} = \frac{6}{5} = 1.2 , \\E(Y) &= 3 \cdot \frac{3}{5} + 4 \cdot \frac{2}{5} = \frac{17}{5} = 3.4 , \\E(Y^2) &= 9 \cdot \frac{3}{5} + 16 \cdot \frac{2}{5} = \frac{59}{5} = 11.8 , \\\sigma^2(Y) &= E(Y^2) - E(Y)^2 = \frac{59}{5} - \left(\frac{17}{5}\right)^2 = \frac{6}{25} = 0.24 .\end{aligned}$$

Der Erwartungswert der Produktvariable XY kann auf mehrere Arten berechnet werden:

$$\begin{aligned}E(XY) &= \sum_{i,j} i \cdot j \cdot P(X = i \text{ und } Y = j) \\&= 0 \cdot 3 \cdot \frac{1}{5} + 1 \cdot 3 \cdot \frac{1}{5} + 2 \cdot 3 \cdot \frac{1}{5} + 0 \cdot 4 \cdot \frac{0}{5} + 1 \cdot 4 \cdot \frac{1}{5} + 2 \cdot 4 \cdot \frac{1}{5} = \frac{21}{5} .\end{aligned}$$

Alternative Rechnung:

$$E(XY) = \sum_k k \cdot P(XY = k) = 0 \cdot \frac{1}{5} + 3 \cdot \frac{1}{5} + 4 \cdot \frac{1}{5} + 6 \cdot \frac{1}{5} + 8 \cdot \frac{1}{5} = \frac{21}{5} .$$

Die Variablen X und Y sind abhängig, z.B. gilt

$$P(X = 0 \text{ und } Y = 3) = \frac{1}{5} \neq P(X = 0) \cdot P(Y = 3) = \frac{3}{25} .$$

Aufgabe 71: (Erwartungswerte, Unabhängigkeit)

Es seien X_1, \dots, X_n unabhängige Zufallsvariable mit dem gemeinsamen Erwartungswert $\mu = E(X_i)$ und der gemeinsamen Streuung $\sigma = \sigma(X_i)$. Setze $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ und $V = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Zeige: $E(\bar{X}) = \mu$, $E(V) = \sigma^2$.

Musterlösung:

Der Erwartungswert des „Mittelwerts“ \bar{X} ist offensichtlich der gemeinsame Erwartungswert:

$$E(\bar{X}) = \frac{1}{n} \sum_i E(X_i) = \frac{1}{n} \sum_i \mu = \mu .$$

Für die „Abweichungssumme“ V gilt:

$$\begin{aligned}(n-1)V &= \sum_i (X_i - \bar{X})^2 = \sum_i (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\&= \left(\sum_i X_i^2\right) - 2\left(\sum_i X_i\right)\bar{X} + n\bar{X}^2 = \left(\sum_i X_i^2\right) - 2n\bar{X}\bar{X} + n\bar{X}^2 \\&= \left(\sum_i X_i^2\right) - n\bar{X}^2 .\end{aligned}$$

Für den Erwartungswert folgt hieraus

$$E\left(\sum_i (X_i - \bar{X})^2\right) = \left(\sum_i E(X_i^2)\right) - nE(\bar{X}^2)$$

und mit

$$E(X_i^2) = E(X_i^2) - E(X_i)^2 + E(X_i)^2 = \sigma^2 + \mu^2$$

ergibt sich

$$E\left(\sum_i (X_i - \bar{X})^2\right) = \left(\sum_i (\sigma^2 + \mu^2)\right) - n E(\bar{X}^2) = n\sigma^2 + n\mu^2 - n E(\bar{X}^2). \quad (\#)$$

Für den letzten Term berechnet man:

$$n E(\bar{X}^2) = \frac{1}{n} E\left(\left(\sum_i X_i\right)^2\right) = \frac{1}{n} E\left(\left(\sum_i X_i\right)\left(\sum_j X_j\right)\right) = \frac{1}{n} \sum_{i,j} E(X_i X_j).$$

Für $i \neq j$ gilt wegen der Unabhängigkeit $E(X_i X_j) = E(X_i)E(X_j) = \mu^2$, während $E(X_i^2) = \sigma^2 + \mu^2$ für die Summanden mit $i = j$ gilt. Es folgt

$$\begin{aligned} n E(\bar{X}^2) &= \frac{1}{n} \sum_i E(X_i^2) + \frac{1}{n} \sum_{\substack{i,j \\ i \neq j}} E(X_i X_j) \\ &= \frac{1}{n} \sum_i (\sigma^2 + \mu^2) + \frac{1}{n} \sum_{\substack{i,j \\ i \neq j}} \mu^2 \\ &= \sigma^2 + \mu^2 + \frac{n(n-1)}{n} \mu^2 = \sigma^2 + n\mu^2. \end{aligned}$$

Aus (#) ergibt sich

$$(n-1) E(V) = E\left(\sum_i (X_i - \bar{X})^2\right) = n\sigma^2 + n\mu^2 - n E(\bar{X}^2) = (n-1)\sigma^2.$$

Aufgabe 72: (Unabhängige Poisson-Variablen)

Zeige, dass die Summe von zwei unabhängigen Poisson-verteilten Zufallsvariablen wieder Poisson-verteilt ist.

Musterlösung:

Seien X_1, X_2 die Zufallsvariablen:

$$P(X_1 = k_1) = \frac{\lambda^{k_1} e^{-\lambda_1}}{k_1!}, \quad P(X_2 = k_2) = \frac{\lambda^{k_2} e^{-\lambda_2}}{k_2!}, \quad k_1, k_2 = 0, 1, 2, \dots$$

Die Summenvariable $Y = X_1 + X_2$ nimmt offensichtlich wieder Werte $k \in \{0, 1, 2, \dots\}$ an:

$$\begin{aligned} P(Y = k) &= P(X_1 + X_2 = k) = \underbrace{\sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty}}_{k_1+k_2=k} P(X_1 = k_1 \text{ und } X_2 = k_2). \\ &= \sum_{k_1=0}^k P(X_1 = k_1 \text{ und } X_2 = k - k_1). \end{aligned}$$

Wegen der Unabhängigkeit von X_1 und X_2 faktorisiert die gemeinsame Verteilung:

$$P(X_1 = k_1 \text{ und } X_2 = k - k_1) = P(X_1 = k_1) \cdot P(X_2 = k - k_1)$$

$$= \frac{\lambda_1^{k_1} \lambda_2^{(k-k_1)}}{k_1! (k-k_1)!} e^{-\lambda_1-\lambda_2} = \lambda_2^k e^{-\lambda_1-\lambda_2} \left(\frac{\lambda_1}{\lambda_2}\right)^{k_1} \frac{1}{k_1! (k-k_1)!}.$$

Es folgt

$$\begin{aligned} P(X_1 + X_2 = k) &= \sum_{k_1=k}^{\infty} \lambda_2^k e^{-\lambda_1-\lambda_2} \left(\frac{\lambda_1}{\lambda_2}\right)^{k_1} \frac{1}{k_1! (k-k_1)!} \\ &= \lambda_2^k e^{-\lambda_1-\lambda_2} \sum_{k_1=0}^k \left(\frac{\lambda_1}{\lambda_2}\right)^{k_1} \frac{1}{k_1! (k-k_1)!}. \end{aligned}$$

Es gilt, die verbleibende Summe zu vereinfachen. Wir erweitern zunächst mit $k!$:

$$\begin{aligned} P(X_1 + X_2 = k) &= \lambda_2^k e^{-\lambda_1-\lambda_2} \frac{1}{k!} \sum_{k_1=0}^k \left(\frac{\lambda_1}{\lambda_2}\right)^{k_1} \frac{k!}{k_1! (k-k_1)!} \\ &= \lambda_2^k e^{-\lambda_1-\lambda_2} \frac{1}{k!} \sum_{k_1=0}^k \binom{k}{k_1} \left(\frac{\lambda_1}{\lambda_2}\right)^{k_1}. \end{aligned}$$

Die verbleibende Summe lässt sich nach Binomi $(a+b)^k = \sum_{i=0}^k \binom{k}{i} a^i b^{k-i}$ vereinfachen:

$$\begin{aligned} P(X_1 + X_2 = k) &= \lambda_2^k e^{-\lambda_1-\lambda_2} \frac{1}{k!} \left(\frac{\lambda_1}{\lambda_2} + 1\right)^k = \lambda_2^k e^{-\lambda_1-\lambda_2} \frac{1}{k!} \frac{(\lambda_1 + \lambda_2)^k}{\lambda_2^k} \\ &= \frac{(\lambda_1 + \lambda_2)^k}{k!} e^{-(\lambda_1 + \lambda_2)}. \end{aligned}$$

Dies ist wieder eine Poisson-Verteilung zum Parameter $\lambda_1 + \lambda_2$.

Anmerkung: Der λ -Parameter der Poisson-Verteilung ist der Erwartungswert. Wegen $E(X_1 + X_2) = E(X_1) + E(X_2)$ war es von vornherein klar, dass der Poisson-Parameter der Summe die Summe der Poisson-Parameter sein muss (sobald man weiß, dass die Summe Poisson-verteilt ist).

Aufgabe 73: (Faltung, Normalverteilung)

Es seien X und Y unabhängige kontinuierliche Zufallsvariablen mit den Dichten ρ_X und ρ_Y .

- a) Zeige, dass die Dichtefunktion $\rho_{X_1+X_2}$ der Variablen $X_1 + X_2$ durch die „Faltung“

$$\rho_{X_1+X_2}(y) = \int_{-\infty}^{\infty} \rho_{X_1}(x) \cdot \rho_{X_2}(x-y) dx$$

gegeben ist. Verwende hierzu (ohne Beweis):

$$P(X_1 + X_2 \leq r) = \int \int_{x_1+x_2 \leq r} \rho_{X_1}(x_1) \cdot \rho_{X_2}(x_2) dx_1 dx_2.$$

- b) Zeige, dass die Summe zweier unabhängiger normalverteilter Zufallsvariablen wieder normalverteilt ist. (Vorsicht: Rechnerie!)

Musterlösung:

a) Wegen der vorausgesetzten Unabhängigkeit der Variablen ist die gemeinsame Verteilungsfunktion

$$F_{X_1, X_2}(r_1, r_2) = P(X_1 \leq r_1, X_2 \leq r_2) = \iint_{x_1 \leq r_1, x_2 \leq r_2} \rho_{X_1}(x_1) \cdot \rho_{X_2}(x_2) dx_1 dx_2.$$

Mit

$$P((X_1, X_2) \in E) = \iint_E \rho_{X_1}(x_1) \cdot \rho_{X_2}(x_2) dx_1 dx_2$$

ergibt sich:

$$P(X_1 + X_2 \leq r) = \iint_{x_1 + x_2 \leq r} \rho_{X_1}(x_1) \cdot \rho_{X_2}(x_2) dx_1 dx_2.$$

Mit der Variablentransformation $(x_1, x_2) \rightarrow (y_1, y_2)$ mit $y_1 = x_1, y_2 = x_1 + x_2$ folgt (die Jakobi-Determinante ist 1):

$$\begin{aligned} P(X_1 + X_2 \leq r) &= \iint_{x_1 + x_2 \leq r} \rho_{X_1}(y_1) \cdot \rho_{X_2}(y_2 - y_1) dy_1 dy_2 \\ &= \int_{y_2 \leq r} \int_{y_1 \in \mathbb{R}} \rho_{X_1}(y_1) \cdot \rho_{X_2}(y_2 - y_1) dy_1 dy_2 = \int_{y_2 \leq r} \rho(y_2) dy_2 \end{aligned}$$

mit

$$\rho(y_2) = \int_{-\infty}^{\infty} \rho_{X_1}(y_1) \cdot \rho_{X_2}(y_2 - y_1) dy_1.$$

b) Mit

$$\rho_{X_i}(x) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}, \quad i = 1, 2$$

ergibt sich aus a) die Dichte

$$\rho(y) = \int_{-\infty}^{\infty} \rho_{X_1}(x) \cdot \rho_{X_2}(y-x) dx = \frac{1}{\sigma_1 \cdot \sigma_2 \cdot 2 \cdot \pi} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \cdot e^{-\frac{(y-x-\mu_2)^2}{2\sigma_2^2}} dx$$

der Summenvariable $X_1 + X_2$. Mit etwas Mühsal stellt man fest, dass die Identität

$$-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-x-\mu_2)^2}{2\sigma_2^2} = -\frac{(y-\mu_1-\mu_2)^2}{2 \cdot (\sigma_1^2 + \sigma_2^2)} - \frac{((\sigma_1^2 + \sigma_2^2) \cdot x - \sigma_1^2 \cdot y + \mu_2 \cdot \sigma_1^2 - \mu_1 \cdot \sigma_2^2)^2}{2 \cdot \sigma_1^2 \cdot \sigma_2^2 \cdot (\sigma_1^2 + \sigma_2^2)}$$

gilt, mit der

$$\rho(y) = \frac{1}{\sigma_1 \cdot \sigma_2 \cdot 2 \cdot \pi} \cdot e^{-\frac{(y-\mu_1-\mu_2)^2}{2 \cdot (\sigma_1^2 + \sigma_2^2)}} \cdot \int_{-\infty}^{\infty} e^{-\frac{((\sigma_1^2 + \sigma_2^2) \cdot x - \sigma_1^2 \cdot y + \mu_2 \cdot \sigma_1^2 - \mu_1 \cdot \sigma_2^2)^2}{2 \cdot \sigma_1^2 \cdot \sigma_2^2 \cdot (\sigma_1^2 + \sigma_2^2)}} dx$$

folgt. Das verbleibende Integral liefert mit der Substitution

$$\xi = \frac{(\sigma_1^2 + \sigma_2^2) \cdot x - \sigma_1^2 \cdot y + \mu_2 \cdot \sigma_1^2 - \mu_1 \cdot \sigma_2^2}{\sqrt{2} \cdot \sigma_1 \cdot \sigma_2 \cdot \sqrt{\sigma_1^2 + \sigma_2^2}}$$

den Wert

$$\int_{-\infty}^{\infty} e^{-\frac{((\sigma_1^2 + \sigma_2^2) \cdot x - \sigma_1^2 \cdot y + \mu_2 \cdot \sigma_1^2 - \mu_1 \cdot \sigma_2^2)^2}{2 \cdot \sigma_1^2 \cdot \sigma_2^2 \cdot (\sigma_1^2 + \sigma_2^2)}} dx = \frac{\sqrt{2} \cdot \sigma_1 \cdot \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \cdot \underbrace{\int_{-\infty}^{\infty} e^{-\xi^2} d\xi}_{=\sqrt{\pi}}.$$

Damit ergibt sich

$$\rho(y) = \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2} \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(y - \mu_1 - \mu_2)^2}{2 \cdot (\sigma_1^2 + \sigma_2^2)}}.$$

Dies ist wieder die Dichte einer Normalverteilung, dessen Parameter durch

$$E(X_1 + X_2) = \mu_1 + \mu_2, \quad \text{Var}(X_1 + X_2) = \sigma_1^2 + \sigma_2^2$$

gegeben sind (bekannterweise addieren sich bei Summen von unabhängigen Variablen die Erwartungswerte und die Varianzen).

Aufgabe 74:** (Normalverteilung. 20 Punkte)

Dies ist eine Online-Aufgabe, die bis zum 4.7.06, 23⁵⁹ Uhr, abzuliefern ist.

In einer Anlage wird Milch in 1-Liter Flaschen abgefüllt. Die Abfüllmenge X variiert dabei etwas: Sie sei normalverteilt mit dem Erwartungswert 1.01 (Liter) und der Standardabweichung 0.01. Die FlaschengöÙe variiert unabhängig davon ebenfalls gemäß einer Normalverteilung: Der Erwartungswert ist 1.06 (Liter), die Standardabweichung ist 0.02. (Erwartungswerte und Standardabweichungen werden vom Aufgabenserver zufällig variiert.)

Mit welcher Wahrscheinlichkeit läuft eine Flasche beim Befüllen über?

Anleitung: Betrachte die Variable „FlaschengöÙe minus Einfüllmenge“. Diese ist nach Aufgabe 73.b) normalverteilt.

Musterlösung:

Die Musterlösung gibt's demnächst. Sei X_1 die FlaschengöÙe, sei X_2 die Einfüllmenge. Die Variable $Y = X_1 - X_2$ ist wiederum normalverteilt mit dem Erwartungswert

$$\mu = E(Y) = E(X_1) - E(X_2) = 1.06 - 1.01 = 0.05$$

und der Varianz

$$\sigma^2 = \text{Var}(Y) = \text{Var}(X_1) + \text{Var}(-X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 0.01^2 + 0.02^2,$$

also

$$\sigma = \sqrt{0.01^2 + 0.02^2} = \frac{\sqrt{5}}{100} \approx 0.02236067977.$$

Es geht um die Wahrscheinlichkeit, dass $Y < 0$ gilt (dann läuft die Flasche über). Mit der Substitution $y = (x - \mu)/\sigma$:

$$\begin{aligned} P(Y \leq 0) &= \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \int_{-\infty}^0 e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}} dx = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{-\mu/\sigma} e^{-\frac{y^2}{2}} dy \\ &= \underbrace{\frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^0 e^{-\frac{y^2}{2}} dy}_{=1/2} - \frac{1}{2} \cdot \underbrace{\frac{1}{\sqrt{2 \cdot \pi}} \int_{-\mu/\sigma}^{\mu/\sigma} e^{-\frac{y^2}{2}} dy}_{\Phi(\mu/\sigma)} = \frac{1}{2} - \frac{1}{2} \cdot \Phi\left(\frac{\mu}{\sigma}\right), \end{aligned}$$

wobei eine Wertetabelle für

$$\Phi(r) = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-r}^r e^{-y^2/2} dy$$

auf Seite 81 des Skripts angegeben ist. Mit

$$\frac{\mu}{\sigma} = \frac{0.05}{\sqrt{5}/100} = \sqrt{5} \approx 2.236$$

findet man aus der Tabelle:

$$P(Y \leq 0) = \frac{1}{2} - \frac{1}{2} \cdot \Phi(2.236) \approx \frac{1 - 0.9747}{2} \approx 0.0127.$$

Alternativ (und einfacher) mit MuPAD:

```
>> stats::normalCDF(1.06 - 1.01, 0.01^2 + 0.02^2)(0.0)
```

0.01267365934

Aufgabe 75: (Bedingte Erwartungswerte)

Mein Lotto„system“ besteht darin, dass ich mir ein zufälliges $k \in \{1, \dots, 49\}$ vorgebe und dann in Zukunft jede Woche k (unabhängige) Lottotips abgebe. Nach wievielen Wochen darf ich das erste Mal mit „6 Richtigen“ rechnen? (Vergleiche das Ergebnis mit dem Ergebnis von Aufgabe 51 auf Blatt 9.)

Musterlösung:

Der Stichprobenraum wird in das Ursachensystem U_k ($k = 1, \dots, 49$) zerlegt: $\Omega = \bigcup_{k=1}^{49} U_k$. Nachdem $U_k =$ „ich habe k gewählt“ festgelegt ist, folgt eine Wiederholung des Bernoulli-Experiments „Lottospiel mit k Tips“ mit der Erfolgswahrscheinlichkeit

$$\begin{aligned} &P(\text{„mindestens einmal 6 Richtige bei } k \text{ Tips“}) \\ &= 1 - P(\text{„keine 6 Richtige bei } k \text{ Tips“}) \\ &= 1 - q^k \\ &= 1 - (1 - p)^k, \end{aligned}$$

wobei

$$\begin{aligned} q &= P(\text{„keine 6 Richtige bei einem Tip“}) = 1 - p, \\ p &= P(\text{„6 Richtige bei einem Tip“}) = 1/\binom{49}{6}. \end{aligned}$$

Nach Aufgabe 47.a) ist der Erwartungswert der Anzahl X der Spiele bis zum ersten Erfolg:

$$E(X | U_k) = \frac{1}{\text{Erfolgswahrscheinlichkeit}_k} = \frac{1}{1 - (1 - p)^k}.$$

Wähle ich heute ein beliebiges $k \in \{1, \dots, 49\}$, so ergibt sich nach Satz 2.59 der Vorlesung

$$E(X) = \sum_{k=1}^{49} E(X | U_k) P(U_k) = \frac{1}{49} \sum_{k=1}^{49} \frac{1}{1 - (1 - p)^k}.$$

Der numerische Wert ist

$$E(X) = 1\,278\,293.991\dots \approx 1\,278\,294 \text{ (Versuche)}.$$

Vergleich: der Erwartungswert für die Anzahl der Versuche des Lotto„systems“ aus Aufgabe 51 (mit ständig wechselndem k) war:

$$\frac{1}{25} \binom{49}{6} = 559\,352.64 \approx 559\,353 \text{ (Versuche)}.$$

Aufgabe 76: (Bedingte Wahrscheinlichkeiten und Erwartungswerte, der Satz von Bayes, Chebyshev, Normalverteilung und all das ...)

Bei einem neuen Automodell stellt der Hersteller fest, dass 10% aller bis jetzt eingebauten Kurbelwellen mit Materialfehlern behaftet sind. Die Konstruktion der Kurbelwellen ist auf eine mittlere Laufleistung von 150 000 km ausgelegt mit einer Streuung von 50 000 km, die fehlerhaften Wellen halten in der Regel nur 20 000 km mit einer Streuung von 10 000 km.

- a) Mit welcher Wahrscheinlichkeit wird ein von mir gekauftes Auto dieses Modells die ersten 50 000 km überstehen?
- b) Nach etwa 50 000 km bleibt mein Wagen mit defekter Kurbelwelle liegen. Sollte ich gerechterweise erwarten, dass der Schaden auf Kulanz behoben wird?

Anmerkungen: a) kann ohne Zusatzannahmen per Formel der totalen W'keit und Chebyshev abgeschätzt werden. Für b) braucht man Zusatzannahmen, z.B., die Verteilung für die Lebensdauer der Kurbelwellen: wähle eine Normalverteilung. Schätze das Verhältnis der Wahrscheinlichkeiten für eine gute/schlechte Kurbelwelle bei dem vorgegebenen Ereignis, dass die Kurbelwelle in einem kleinen Intervall um die Laufleistung 50 000 km ihren Dienst einstellt.

Musterlösung:

Sei X die Laufleistung (in km) über einem unbekanntem Stichprobenraum, der sich zerlegt in $\Omega = \Omega_+ \cup \Omega_-$ ($\Omega_{\pm} \hat{=}$ gute/schlechte Kurbelwelle). Gegeben sind die bedingten Erwartungswerte/Streuungen:

$$\begin{aligned}\mu_+ &= E(X | +) = 150\,000, & \sigma_+ &= \sigma(X | +) = 50\,000, \\ \mu_- &= E(X | -) = 20\,000, & \sigma_- &= \sigma(X | -) = 10\,000.\end{aligned}$$

a) Mit der „Formel der totalen Wahrscheinlichkeit“ gilt:

$$\begin{aligned}P(X \geq 50\,000) &= P(X \geq 50\,000 | +) P(+) + P(X \geq 50\,000 | -) P(-) \\ &= 0.9 P(X \geq 50\,000 | +) + 0.1 P(X \geq 50\,000 | -). \quad (\#)\end{aligned}$$

Nach Chebyshev gilt die Abschätzung:

$$\begin{aligned}P(X \geq 50\,000 | +) &= P(X - \mu_+ \geq -100\,000 | +) \\ &\geq P(100\,000 \geq X - \mu_+ \geq -100\,000 | +) \\ &= P(|X - \mu_+| \leq 100\,000 | +) \geq 1 - \frac{\sigma^2(X | +)}{100\,000^2} = 1 - \frac{1}{4} = 0.75.\end{aligned}$$

Für die andere bedingte Wahrscheinlichkeit erhält man per Chebyshev keine sinnvolle Abschätzung nach unten, wir können höchstens folgern, dass dieser Wert nicht allzu groß ist:

$$\begin{aligned}P(X \geq 50\,000 | -) &= P(X - \mu_- \geq 30\,000 | -) \\ &\leq P(X - \mu_- \geq 30\,000 \text{ oder } X - \mu_- \leq -30\,000 | -) \\ &= P(|X - \mu_-| \geq 30\,000 | -) \leq \frac{\sigma^2(X | -)}{30\,000^2} = \left(\frac{1}{3}\right)^2 = 0.1.\end{aligned}$$

Damit können wir in (#) getrost den zweiten Term vernachlässigen:

$$P(X \geq 50\,000) \geq 0.9 P(X \geq 50\,000 | +) \geq 0.9 \cdot 0.75 = 0.675 .$$

Setzt man für die Kurbelwellen jeweils eine $N(\mu_{\pm}, \sigma_{\pm}^2)$ -Verteilung voraus, so ergeben sich konkretere Werte:

$$P(X \geq 50\,000 | \pm) = \frac{1}{\sigma_{\pm} \sqrt{2\pi}} \int_{50\,000}^{\infty} e^{-\frac{(x-\mu_{\pm})^2}{2\sigma_{\pm}^2}} dx .$$

Die Substitution $y = (x - \mu_{\pm})/\sigma_{\pm}$ liefert:

$$\begin{aligned} P(X \geq 50\,000 | \pm) &= \frac{1}{\sqrt{2\pi}} \int_{\frac{50\,000-\mu_{\pm}}{\sigma_{\pm}}}^{\infty} e^{-y^2/2} dy \\ &= \frac{1}{2} \left(1 - \operatorname{sign}\left(\frac{50\,000 - \mu_{\pm}}{\sigma_{\pm}}\right) \Phi\left(\left|\frac{50\,000 - \mu_{\pm}}{\sigma_{\pm}}\right|\right) \right) , \end{aligned}$$

also

$$P(X \geq 50\,000 | +) = \frac{1 + \Phi(2)}{2} \approx 0.977 ,$$

$$P(X \geq 50\,000 | -) = \frac{1 - \Phi(3)}{2} \approx 0.00135$$

und damit

$$P(X \geq 50\,000) \approx 0.9 \cdot 0.977 + 0.1 \cdot 0.00135 = 0.879 .$$

b) Das Ereignis $A =$ „Kurbelwellendefekt bei genau 50 000 km“ hat für eine kontinuierliche Variable die Wahrscheinlichkeit 0, man sollte also nach der Wahrscheinlichkeit für das Ereignis

$$A = \text{„Defekt bei einer Laufleistung im Bereich von } 50\,000 - \epsilon \text{ bis } 50\,000 + \epsilon\text{“}$$

(mit kleinem ϵ) fragen. Nach Bayes gilt

$$P(\pm | A) = \frac{P(A | \pm) P(\pm)}{P(A)} .$$

Die Kulanzfrage sollte man danach messen, mit welcher Wahrscheinlichkeit man sagen kann, ob der Defekt auf den Werksfehler zurückgeht. Die Frage ist also $P(- | A) > P(+ | A)$? Mit Bayes folgt

$$\frac{P(+ | A)}{P(- | A)} = \frac{P(A | +) P(+)}{P(A | -) P(-)} = \frac{P(A | +) \cdot 0.9}{P(A | -) \cdot 0.1} .$$

Um die benötigten $P(A | \pm)$ berechnen zu können, nehmen wir an (da wir nichts Besseres wissen), dass die Lebensdauer der Kurbelwellen $N(\mu_{\pm}, \sigma_{\pm}^2)$ -verteilt ist. Für sehr kleines ϵ ergibt sich

$$P(A | \pm) = \frac{1}{\sigma_{\pm} \sqrt{2\pi}} \int_{50\,000-\epsilon}^{50\,000+\epsilon} e^{-\frac{(x-\mu_{\pm})^2}{2\sigma_{\pm}^2}} dx \approx \frac{2\epsilon}{\sigma_{\pm} \sqrt{2\pi}} e^{-\frac{(50\,000-\mu_{\pm})^2}{2\sigma_{\pm}^2}} .$$

Im Quotienten fällt ϵ heraus:

$$\frac{P(+ | A)}{P(- | A)} = 9 \frac{P(A | +)}{P(A | -)} = 9 \frac{\sigma_- e^{-\frac{(50\,000-\mu_+)^2}{2\sigma_+^2}}}{\sigma_+ e^{-\frac{(50\,000-\mu_-)^2}{2\sigma_-^2}}} = \frac{9}{5} \frac{e^{-2}}{e^{-9/2}} \approx 21.9 .$$

Die Wahrscheinlichkeit, dass es sich um einen „üblichen“ Kurbelwellenschaden handelt, ist damit wesentlich größer als die Wahrscheinlichkeit, dass in meinen Wagen eine schlechte Kurbelwelle eingebaut worden war.
