Zusammenfassung Wahrscheinlichkeit und Statistik

Nicolas Trüssel

7. Januar 2016

1 Diskrete Verteilungen

1.1 Bernoulli Verteilung

Sei $X \sim Be(p)$ (Erfolg 1, Misserfolg 0). Dann ist:

$$P[X = x] = p^x(1-p)^x, x \in \{0, 1\}, \qquad E(X) = p, \qquad \text{Var}(X) = p(1-p), \qquad m_k = p$$

1.2 Binomialverteilung

Sei $X \sim Bin(n, p)$ (Anzahl Erfolge). Dann ist:

$$P[X = k] = \binom{n}{k} p^k(1-p)^{n-k}, k \in \{0, \ldots, n\}, \qquad E(X) = np, \qquad \text{Var}(X) = np(1-p)$$

1.3 Geometrische Verteilung

Sei $X \sim Geom(p)$ (Wartezeit auf ersten Erfolg). Dann ist:

$$P[X = k] = p(1-p)^{k-1}, k \in \{1, 2, 3, \ldots\}, \qquad E(X) = \frac{1}{p}, \qquad \text{Var}(X) = \frac{1-p}{p^2}$$

1.4 Negativ-Binomialverteilung

Sei $X \sim NB(r, p)$ (Wartezeit auf r-ten Erfolg). Dann ist:

$$P[X = k] = \binom{k-1}{r-1} p^r(1-p)^{k-r}, k \in \{r, r+1, \ldots\}, \qquad E(X) = \frac{r}{p}, \qquad \text{Var}(X) = \frac{r(1-p)}{p^2}$$

1.5 Hypergeometrische Verteilung

Sei $X \sim HGeom(n, m, r)$. In einer Urne sind r Kugeln mit einer Eigenschaft, $n-r$ Kugeln ohne diese Eigenschaft. Man zieht m Kugeln ohne Zurücklegen. X beschreibt die Anzahl der Kugeln mit der Eigenschaft. Dann ist:

$$P[X = k] = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}, k \in \{0, 1 \ldots \min(m, r)\}, \qquad E(X) = m\frac{r}{n}, \qquad \text{Var}(X) = m\frac{r}{n}\left(1 - \frac{r}{n}\right)\frac{n-m}{n-1}$$

1.6 Poisson Verteilung

Sei $X \sim Poi(\lambda)$ (für seltene Ereignisse benutzt). Dann gilt:

$$P[X = k] = \boldsymbol{e}^{-\lambda}\frac{\lambda^k}{k!}, k \in \{0, 1, 2, \ldots\}, \qquad E(X) = \lambda, \qquad \text{Var}(X) = \lambda$$

1.6.1 Summe poisson-verteilter Zufalllsvariablen

Sind $X_1 \sim Poi(\lambda_1), X_2 \sim Poi(\lambda_2)$ unabhängig, so ist $X_1 + X_2 = X \sim Poi(\lambda_1 + \lambda_2)$.

1.7 Beziehung zur Binomialverteilung

Für grosse n und kleine p kann man die Binomial-Wahrscheinlichkeit approximativ durch die Poisson-Wahrschinlichkeit berechnen wobei $\lambda = np$.

2 Stetige Verteilungen

2.1 Gleichverteilung

Sei $X \sim \mathcal{U}(a,b)$. Dann ist

$$f_X(t) = \begin{cases} \frac{1}{b-a} & a \leq t \leq b \\ 0 & \text{sonst} \end{cases} \qquad F_X(t) = \begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ 1 & t > b \end{cases} \qquad E(X) = \frac{a+b}{2} \qquad \operatorname{Var}(X) = \frac{1}{12}(b-a)^2$$

2.2 Exponentialverteilung

Sei $X \sim Exp(\lambda)$ (für Wartezeiten und Lebensdauern). Dann gilt:

$$f_X(t) = \begin{cases} \lambda \boldsymbol{e}^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad F_X(t) = \begin{cases} 1 - \boldsymbol{e}^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad E(X) = \frac{1}{\lambda} \qquad \operatorname{Var}(X) = \frac{1}{\lambda^2}$$

2.3 Normalverteilung

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann ist:

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \boldsymbol{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}, t \in \mathbb{R}, \qquad E(X) = \mu, \qquad \operatorname{Var}(X) = \sigma^2$$

2.3.1 Standardnormalverteilung

Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $\dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$. Also ist $F_X(t) = \Phi\left(\dfrac{t-\mu}{\sigma}\right)$

2.3.2 Summe

Sind $X \sim \mathcal{N}(\mu_X, \sigma_X^2), Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ unabhängig, so sind ist $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

2.4 Pareto-Verteilung

Sei $X \sim Par(\alpha, x_0) \quad \alpha, x_0 > 0$ (für Katastrophen). Dann ist:

$$f_X(x) = \begin{cases} \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} & x \geq x_0 \\ 0 & x < x_0 \end{cases} \qquad F_X(x) = 1 - \left(\frac{x_0}{x}\right)^{\alpha}, x \geq x_0 \qquad E(X) = \begin{cases} x_0 \frac{\alpha}{\alpha-1} & \alpha > 1 \\ \infty & \alpha < 1 \end{cases}$$

3 Definitionen und Umformungen

3.1 Bedingte Wahrscheinlichkeit

3.1.1 Definition

$$P[A|B] := \frac{P[A \cap B]}{P[B]}$$

3.1.2 Umformungen

$$P[A^C|B] = 1 - P[A|B]$$

Sind A und B unabhängig, so gilt:

$$P[A \cap B|C] = P[A|C]P[B|C]$$

3.2 Formel von Bayes

Sei $A_1, \ldots A_n$ eine Zerlegung des Grundraumes mit $P[A_i] > 0$ und B ein Ereignis mit $P[B] > 0$. Dann gilt für jedes k:

$$P[A_k|B] = \frac{P[B|A_k]P[A_k]}{\sum_{i=1}^{n} P[B|A_i]P[A_i] = P[B]}$$

3.3 Erwartungswert

3.3.1 Definition

$$E(X) = \sum_{x \in \mathcal{W}(X)} xP[X = x], \qquad E(X) = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x$$

3.3.2 Umformungen

Sei $Y = g(X)$:

$$E(Y) = \sum_{x \in \mathcal{W}(X)} g(x)P[X = x], \qquad E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$$

Insbesondere gilt:

$$E(aX + b) = aE(X) + b, \qquad E(aX + bY) = aE(X) + bE(y)$$

3.3.3 Summe von Zufallsvariablen

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E\left(X_i\right)$$

3.3.4 Produkt von Zufallsvariablen

Sind $X_i \ldots X_n$ unabhängig, so ist

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E\left(X_i\right)$$

3.4 Varianz

3.4.1 Definition

$$\mathrm{Var}(X) = E\left((X - E(X))^2\right)$$

3.4.2 Umformungen

Falls $E(X^2) < \infty$:

$$\mathrm{Var}(X) = E(X^2) - E(X)^2, \qquad \mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$$

3.4.3 Summe von Zufallsvariablen

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

3.5 Kovarianz

3.5.1 Definition

$$\mathrm{Cov}(X, Y) = E\left[(X - E\left(X\right)) \cdot (Y - E\left(Y\right))\right]$$

3.5.2 Umformungen

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

Die Kovarianz ist eine positiv semidefinite Bilinearform, d.h es gilt:

$$\begin{aligned}
\mathrm{Cov}(aX+b, cY+d) &= ac\,\mathrm{Cov}(X,Y) \\
\mathrm{Cov}(X, (aY+b)+(cZ+d)) &= a\,\mathrm{Cov}(X,Y) + c\,\mathrm{Cov}(X,Z) \\
\mathrm{Cov}(X,Y) &= \mathrm{Cov}(Y,X) \\
\mathrm{Cov}(X,X) &\geq 0
\end{aligned}$$

3.6 Randverteilung

Haben X und Y die gemeinsame Verteilungsfunktion F, so ist $F_X : \mathbb{R} \to [0,1]$,

$$x \mapsto F_X(x) := \lim_{y\to\infty} F(x,y)$$

die Randverteilung von X.

Wenn X und X die gemeinsame Dichte f (p im diskreten Fall) haben, so haben auch die Randverteilungen Dichten:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y, \qquad p_X(x) = \sum_{y\in\mathcal{W}(Y)} p(x,y)$$

3.7 Zentraler Grenzwertsatz

Sei $X_1, X_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit $E(X_i) = \mu, \mathrm{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$. Dann ist:

$$E(S_n) = n\mu, \quad \mathrm{Var}(S_n) = n\sigma^2, \quad S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1), \quad \text{für } n \text{ gross}$$

4 Ungleichungen

4.1 Markov-Ungleichung

Sei X eine Zufallsvariable und $g : \mathcal{W}(X) \to [0,\infty)$ eine wachsende Funktion. Für alle $c \in \mathbb{R}$ mit $g(c) > 0$ gilt:

$$P[X \geq c] \leq \frac{E(g(X))}{g(c)}$$

4.2 Chebyshev-Ungleichung

Sei X eine Zufallsvariable mit endlicher Varianz. Für jedes $b > 0$ gilt:

$$P[|X - E(X)| \geq b] \leq \frac{\mathrm{Var}(X)}{b^2}$$

4.3 Chernoff-Schranken

Seien $X_1, \ldots, X_n \sim Be(p_i)$ unabhängig und $S_n = \sum\limits_{i=1}^{n} X_i$, $\mu_n = E[S_n] = \sum\limits_{i=1}^{n} p_i$ und $\delta > 0$. Dann gilt:

$$P[S_n \geq (1+\delta)\mu_n] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mu_n}$$

4.3.1 Allgemeinere Ungleichung

Seien $X_1, \dots, X_n$ i.i.d. Zufallsvariablen, für welche die momenterzeugende Funktion $M_X(t)$ für alle $t \in \mathbb{R}$ endlich ist ($M_X(t) := E(\boldsymbol{e}^{tX})$). Für jedes $b \in \mathbb{R}$ gilt:

$$P[S_n \geq b] \leq \exp\left(\inf_{t\in\mathbb{R}} (n \log M_X(t) - tb)\right)$$

4.4 Schwaches Gesetz der Grossen Zahlen

Sei $X_1, X_2, \dots$ eine Folge von unabhängigen Zufallsvariablen mit dem selben Erwartungswert μ und der gleichen Varianz σ^2. Dann gilt:

$$P\left[\left|\overline{X}_n - \mu\right| > \varepsilon\right] \xrightarrow[n\to\infty]{} 0, \quad \text{für jedes } \varepsilon > 0$$

Das heisst, dass mit beliebig grosser Wahrscheinlichkeit der Wert von $\overline{X}_n$ für hinreichend grosse n beliebig nahe bei μ liegt.

4.5 Starkes Gesetz der Grossen Zahlen

Sei $X_1, X_2, \dots$ eine Folge von unabhängigen Zufallsvariablen mit der selben Verteilung und endlichem Erwartungswert μ. Es gilt:

$$\overline{X}_n \xrightarrow[n\to\infty]{} \mu, \quad \text{P-fastsicher,} \qquad \text{d.h.} \qquad P\left[\left\{\omega \in \Omega | \overline{X}_n(\omega) \xrightarrow[n\to\infty]{} \mu\right\}\right] = 1$$

5 Schätzer

5.1 Eigenschaften

5.1.1 Erwartungstreu

Ein Schätzer T heisst erwartungstreu für ϑ, falls gilt $E_\vartheta[T] - \vartheta = 0$, die linke Seite dieser Gleichung heisst Bias.

5.1.2 Mean Square Error

Als Mean Square Error bezeichnet man

$$MSE_\vartheta(T) := E_\vartheta\left((T - \vartheta)^2\right)$$

5.1.3 Konsistenz

Eine Folge von Schätzern $T^{(n)}$ ist konsistent für ϑ, falls

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P_\vartheta\left[\left|T^{(n)} - \vartheta\right| > \varepsilon\right] = 0$$

5.2 Maximum Likelihood

Die Likelihood-Funktion ist:

$$L(x_1, \dots x_n; \vartheta) := \begin{cases} p(x_1, \dots x_n; \vartheta) & \text{im diskreten Fall} \\ f(x_1, \dots x_n; \vartheta) & \text{im stetigen Fall} \end{cases}$$

Um den Paramteter ϑ zu schätzen, suche das Maximum der Likelihood-Funktion (meist zuerst den Logarithmus der ML-Funktion bestimmen und dann ableiten).

5.2.1 Bernoulli-Verteilung

Seien $X_1, \dots, X_n \overset{i.i.d.}{\sim} Be(p)$ der ML-Schätzer für $\vartheta = p$ ist

$$\hat{\vartheta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$$

5.2.2 Normalverteilung

Seien $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$ (μ und σ sind unbekannt). Dann ist der ML-Schätzer für $\vartheta = \begin{pmatrix} \mu & \sigma^2 \end{pmatrix}^T$:

$$\hat{\vartheta} = \begin{pmatrix} \overline{X}_n \\ \frac{1}{n} \sum\limits_{i=1}^{n} \left(X_i^2\right) - \overline{X}_n^{\ 2} \end{pmatrix}$$

Da der Schätzer für die Varianz nicht erwartunstreu ist wählt man für σ^2 meist den Schätzer

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$$

5.3 Momentenschätzer

5.3.1 Definition

Sei $(X,)X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(\,\cdot\,; \vartheta), \vartheta \in \Theta \subset \mathbb{R}^m$ Das k-te Moment ist definiert als:

$$m_k := E_\vartheta \left(X^k\right) = m_k(\vartheta)$$

Um die m_k zu schätzen kann man die Momentenschätzer verwenden:

$$\hat{m}_{k,n} = \frac{1}{n} \sum_{i=1}^{k} x_i^k$$

Um daraus ϑ zu erhalten setze $\hat{m}_{k,n} = m_k(\vartheta)$ und löse nach ϑ auf. Dazu müssen die Parameter als Funktion der Momente formuliert werden, zum Beispiel $\mathrm{Var}(X) = E(X^2) - E(X)^2 = m_2 - m_1^2$

6 Tests

6.1 Bestimmung des kritischen Bereiches

Für ein gegebenes Signifikanzniveau α, eine Teststatistik T und Nullhypothese Θ_0 kann der kritische Bereich K (bei bekannter Form) so bestimmt werden:

$$\sup_{\vartheta \in \Theta_0} P_\vartheta[T \in K] \leq \alpha$$

6.2 Likelihood-Quotient

Für feste ϑ_0 und ϑ_A ist der Likelihood-Quotient ideal um eine Teststatistik zu finden:

$$R(x_1, \ldots, x_n; \vartheta_0, \vartheta_A) := \frac{L(x_1, \ldots, x_n; \vartheta_0}{L(x_1, \ldots, x_n; \vartheta_A)}$$

Der kritische Bereich hat die Form $[0, c)$. Mit Hilfe des Supremums für Zähler und Nenner, kann diese Methode auch für kompliziertere Hypothesen ausgeweitet werden.

6.3 Z-Test

Seien $X_1, \ldots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$ mit σ bekannt und $H_0 : \vartheta = \vartheta_0$. Dann ist eine geeignete Teststatistik

$$Z := \frac{\overline{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1), \quad \text{unter } P_{\vartheta_0}$$

Die Form des kritischen Bereiches K hängt von der Alternativhypothese ab: Für $H_A : \vartheta > \vartheta_0$ ist $K = (c, \infty)$, für $H_A : \vartheta < \vartheta_0$ ist $K = (-\infty, c)$ und für $H_A : \vartheta \neq \vartheta_0$ ist $K = (-\infty, -c) \cup (c, \infty)$.

6.4 T-Test

Selbe Voraussetzungen wie Z-Test, jedoch ist σ unbekannt.
Zuerst wird die Varianz mit dem Schätzer $S^2 := \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2$ bestimmt. Die Teststatistik ist

$$T := \frac{\overline{X}_n - \vartheta_0}{S/\sqrt{n}} \sim t_{n-1}, \quad \text{unter } P_{\vartheta_0}$$

6.5 Gepaarte Stichproben

Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_Y, \sigma^2)$ Daten von gepaarten Stichproben mit gleicher Varianz σ . Wir definieren $\Delta_i := X_i - Y_i$

6.5.1 Bekannte Varianz

Ist die Varianz bekannt, so sind $\Delta_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X - \mu_Y, 2\sigma^2)$ und die Teststatistik ist

$$\frac{(\overline{X}_n - \overline{Y}_n) - (\mu_X - \mu_Y)}{\sqrt{2}\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

6.5.2 Unbekante Varianz

Ist die Varianz unbekannt, so schätzt man zunächst die Varianz als

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(\Delta_i - \overline{\Delta}_i\right)^2$$

Dann ist

$$\frac{(\overline{X}_n - \overline{Y}_n) - (\mu_X - \mu_Y)}{S/\sqrt{n}} \sim t_{n-1}$$

6.6 Ungepaarte Stichproben

6.6.1 Bekannte Varianz

Seien $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ mit bekannter und identischer Varianz σ^2. Dann machen wir einen Z-Test mit der Teststatistik:

$$\frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1)$$

6.6.2 Unbekannte Varianz

Seien $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ und $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$ mit identischer, aber unbekannter Varianz σ^2. Wir berechnen zuerst die beiden empirischen Varianzen:

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \qquad \text{und} \qquad S_Y^2 = \frac{1}{m-1}\sum_{i=1}^m (Y_i - \overline{Y}_n)^2$$

Mit

$$S^2 = \frac{1}{m+n-2}\left((n-1)\,S_X^2 + (m-1)\,S_Y^2\right)$$

Ist die Teststatistik

$$\frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

7 Satz 7.1

Seien $X_1, \ldots, X_m \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

1. $\overline{X}_n \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$ und $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.

2. $\dfrac{n-1}{\sigma^2} S^2 = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ ist χ^2-verteilt mit $n-1$ Freiheitsgraden.

3. $\overline{X}_n$ und S^2 sind unabhängig.

4. Der Quotient
$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{S/\sigma} = \frac{\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{1}{n-1}\dfrac{n-1}{\sigma^2} S^2}}$$
ist t-verteilt mit $n-1$ Freiheitsgraden.