SS 2014

Zentralübung zur Vorlesung Diskrete Wahrscheinlichkeitstheorie

Dr. Werner Meixner

Fakultät für Informatik
TU München

http://www14.in.tum.de/lehre/2014SS/dwt/uebung/

15. Mai 2014

ZÜ IV

Übersicht:

1. Übungsbetrieb Fragen
2. Thema Poisson-Verteilung
3. Vorbereitung TA Blatt 5

1. Übungsbetrieb

Fragen?

Vorschläge?

2. Thema: Poisson-Verteilung

Die Poisson-Verteilung ist Grenzwert der Binomialverteilung.

Sei λ eine positive reelle Zahl.
Dann kann man eine Folge von p_n's betrachten mit $p_n = \frac{\lambda}{n}$ und für jedes k mit $0 \le k \le n$ den folgenden Grenzübergang beweisen.

$$\begin{aligned}
\binom{n}{k} p_n^k (1-p_n)^{n-k} &= \frac{(n \cdot p_n)^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \cdot (1-p_n)^{-k} \cdot (1-p_n)^n \\
&= \frac{\lambda^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^n
\end{aligned}$$

$\Rightarrow$

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}\,.$$

Daraus folgt

$$\lim_{n\to\infty} b(k;n,p_n) \;=\; e^{-\lambda}\cdot\frac{\lambda^k}{k!}\,.$$

Die Wahrscheinlichkeit $b(k;n,p_n)$ konvergiert also für $n\to\infty$ gegen die Wahrscheinlichkeit, dass eine Poisson-verteilte Zufallsvariable mit Parameter λ den Wert k annimmt.

Insofern steht die Poisson Verteilung in Zusammenhang mit den Matrizes der Binomialverteilung der ZÜ III und insbesondere (bekanntlich) mit der Binomialverteilung.

3. Vorbereitung der TA Blatt 5

3.1 VA 1

X sei Poisson-verteilt.

Berechnen Sie $\mathbb{E}[(X+1)^{-1}]$.

Lösung

X ist eine Zufallsvariable, d.h.,
X ist eine Abbildung $X : \Omega \to \mathbb{R}$,
wobei ein Wahrscheinlichkeitsraum $(\Omega, \Pr)$ vorausgesetzt wird.

X ist Poisson-verteilt, d.h. nach Voraussetzung gilt
$X \sim \mathrm{Po}(\lambda)$ für irgendein bestimmtes $\lambda \geq 0$.

Die diskrete Dichtefunktion von X ist dann $f_X : \mathbb{R} \to [0, 1]$ mit
$f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$ für $i \in \mathbb{N}_0$ und $f_X(i) = 0$ sonst.

$$
\begin{aligned}
\mathbb{E}[(X+1)^{-1}] &= \sum_{i\in W_X} \frac{1}{i+1} f_X(i) \\
&= \sum_{i=0}^{\infty} \frac{1}{i+1} \cdot \frac{e^{-\lambda}\lambda^i}{i!} \\
&= e^{-\lambda}\lambda^{-1} \sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{(i+1)!} \\
&= e^{-\lambda}\lambda^{-1} \left(-1 + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}\right) \\
&= e^{-\lambda}\lambda^{-1}(-1+e^{\lambda}) \\
&= \frac{1-e^{-\lambda}}{\lambda}\,.
\end{aligned}
$$

Bemerkung: Für eine Poisson-verteilte Variable X können systematisch alle Erwartungswerte für beliebige Polynome in X in geschlossener Form berechnet werden.

Beispiel: $X^{\underline{n}}$.

Die vorliegende Aufgabe zeigt, dass diese Methodik sogar auf die Funktion $(X+1)^{-1}$ anwendbar ist.

3.2 VA 2

Eine Firma stellt Kuchen mit Rosinen her. Hierfür werden $\lambda \cdot N$ Rosinen in den Teig für N Kuchen gegeben, woraufhin die ganze Masse so gut durchmischt wird, dass wir annehmen können, dass jede einzelne Rosine mit derselben Wahrscheinlichkeit in einem der Kuchen landet.

N ist unbekannt und groß.

Wie groß muss die durchschnittliche Zahl λ von Rosinen pro Kuchen sein, wenn höchstens durchschnittlich jeder hundertste Kuchen keine Rosinen enthalten darf?

Lösung

Aus der Problemstellung entnimmt man, dass die durchschnittliche Anzahl der Rosinen pro Kuchen mit λ gegeben ist.

Über N bzw. die Anzahl $R = \lambda \cdot N$ der Rosinen ist nur bekannt, dass diese Zahlen sehr groß sind.

Wir betrachten das Problem aus der Sicht eines einzelnen Kuchens und stellen uns vor, dass dem Kuchen eine (potentiell unendliche) Anzahl X von Rosinen zugeteilt werden, wobei die einzelnen Rosinen immer mit gleicher Wahrscheinlichkeit zugeteilt werden.

Wenn jede Rosine die Laplace-verteilte Wahl hat, auf einem von N Kuchen zu landen, dann wird die Rosine auf einem einzelnen, betrachteten Kuchen mit Wahrscheinlichkeit $\frac{1}{N}$ landen.

Wir betrachten für jede der R Rosinen eine Bernoulli-verteilte Zufallsvariable X_i.

$X_i = 1$ bedeutet, dass die Rosine i auf dem betrachteten Kuchen von insgesamt N Kuchen landet.

Die Anzahl der Rosinen, die auf dem Kuchen landen ist dann durch eine binomialverteilte Zufallsvariable X gegeben:

$$X = \sum_{i=0}^{R} X_i \,.$$

Der Erwartungswert von X ist $\mathbb{E}[X] = \frac{R}{N}$.

Wenn der Erwartungswert mit λ gegeben sein soll, dann folgt

$$\lambda = \frac{R}{N} \quad \text{und} \quad R = \lambda \cdot N\,.$$

Für die Wahrscheinlichkeit $\Pr[X = k]$, dass genau k Rosinen auf dem betrachteten Kuchen landen, gilt mit $p_N = \frac{1}{N}$

$$\Pr[X = k] = \binom{R}{k} p_N^k (1 - p_N)^{R-k} = b(k; R, p_N)$$

und approximativ

$$\lim_{R \to \infty} b(k; R, p_N) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}\,.$$

Die Verteilung der Rosinen auf den Kuchen, d. h. der Zufallsvariablen X, nehmen wir also approximativ als Poisson-Verteilung an, d. h.

$f_X : \mathbb{R} \to [0,1]$ mit $f_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$ für $i \in \mathbb{N}_0$ und $f_X(i) = 0$ sonst. Dann erhalten wir

$$\Pr[X = 0] = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}\,.$$

Wenn durchschnittlich höchstens jeder hundertste Kuchen frei von Rosinen sein darf, dann bedeutet dies, dass $\Pr[X = 0] \leq \frac{1}{100}$ erfüllt werden muss. Es folgt

$$\lambda \geq \ln 100 \approx 4,605\,.$$

3.3 VA 3

Zwei Arbeiter A und B kontrollieren unabhängig eine Tagesproduktion. A und B protokollieren k_1 bzw. k_2 tatsächliche Fehler. Es sei n die Anzahl der tatsächlich aufgetretenen Produktionsfehler. Wir nehmen an, dass die Arbeiter jeden der n Fehler mit Wahrscheinlichkeit p_1 bzw. p_2 registriert haben.

Es seien X_1 bzw. X_2 die Zufallsvariablen, die die Anzahl der von Arbeiter A bzw. B gefundenen Fehler angeben. Wie sind die X_i verteilt? Für welche Werte von n gilt

$$\Pr[|\tfrac{1}{n}X_i - p_i| \geq 0.1] \leq 0.01\,?$$

Bemerkung: Man möchte z.B. eine zuverlässige Entscheidung, welche der beiden p_i größer ist.

Lösung

Es gilt $X_i \sim \mathrm{Bin}(n; p_i)$.
Daraus folgt $\mathbb{E}[X_i] = n \cdot p_i$, d. h. $\mathbb{E}[\frac{1}{n} X_i] = p_i$.

Wir wenden die Chebyshev-Ungleichung wie folgt an.

$$\Pr[|\tfrac{1}{n} X_i - p_i| \geq 0.1] \leq \frac{\mathrm{Var}[\frac{1}{n} X_i]}{0.01} \,.$$

Weiter gilt $\mathrm{Var}[\frac{1}{n} X_i] = \frac{1}{n^2} \cdot n \cdot p_i(1 - p_i)$.

Wegen $p_i \in [0, 1]$ gilt $p_i(1 - p_i) \leq \frac{1}{4}$ und es folgt

$$\Pr[|\tfrac{1}{n} X_i - p_i| \geq 0.1] \leq \frac{\mathrm{Var}[\frac{1}{n} X_i]}{0.01} \leq \frac{1}{0{,}04n} \,.$$

Damit haben wir die hinreichende Bedingung

$$n \geq 2500\,.$$

Bemerkung: Solange wir keine weiteren Informationen zu den Wahrscheinlichkeiten p_1 und p_2 haben, können wir die Abschätzung nicht verbessern.

Anwendung: Falls z.B. $n = 3000$ bekannt ist, und

$$X_1 + 0,1 \cdot n \leq X_2 - 0,1 \cdot n$$

gilt, dann ist mit 99 Prozent Sicherheit Arbeiter B besser zur Fehlersuche geeignet als Arbeiter A.