Script generated by TTT

Title: Esparza: DWT (21.06.2012)

Date: Thu Jun 21 14:18:08 CEST 2012

Duration: 87:17 min

Pages: 29

Wegen $\lim_{n\to\infty}(1-\frac{\lambda}{n})^n = e^{-\lambda}$ gilt daher für die Zufallsvariablen $Y_n := \frac{1}{n}X_n$, dass

$$\begin{aligned}\lim_{n\to\infty}\Pr[Y_n \le t] &= \lim_{n\to\infty}\Pr[X_n \le t\cdot n]\\ &= \lim_{n\to\infty}\left[1-\left(1-\frac{\lambda}{n}\right)^{tn}\right]\\ &= 1-e^{-\lambda t}.\end{aligned}$$

Die Folge Y_n der (skalierten) geometrisch verteilten Zufallsvariablen geht also für $n\to\infty$ in eine exponentialverteilte Zufallsvariable mit Parameter λ über.

Beispiel 87

Über das Cäsium-Isotop $^{134}_{55}$Cs ist bekannt, dass es eine mittlere Lebensdauer von ungefähr 3,03 Jahren oder $1{,}55\cdot 10^6$ Minuten besitzt. Die Zufallsvariable X messe die Lebenszeit eines bestimmten $^{134}_{55}$Cs-Atoms. X ist exponentialverteilt mit dem Parameter

$$\lambda = \frac{1}{\mathbb{E}[X]} = \frac{1}{1{,}55\cdot 10^6} \approx 0{,}645\cdot 10^{-6}\left[\frac{1}{\text{min}}\right]$$

Da λ den Kehrwert einer Zeit als Einheit besitzt, spricht man von der Zerfallsrate. Auch bei anderen Anwendungen ist es üblich, λ als Rate einzuführen.

4.4 Minimum von Exponentialverteilungen

Warten auf mehrere Ereignisse

Satz 88

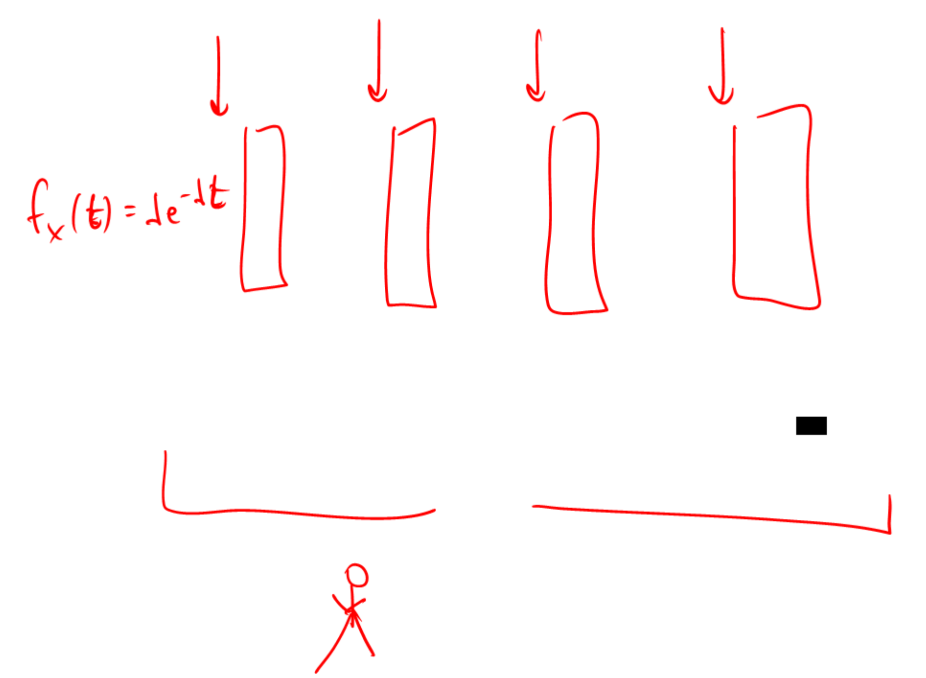
Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und exponentialverteilt mit den Parametern $\lambda_1, \ldots, \lambda_n$. Dann ist auch $X := \min\{X_1, \ldots, X_n\}$ exponentialverteilt mit dem Parameter $\lambda_1 + \ldots + \lambda_n$.

Beweis:

Der allgemeine Fall folgt mittels Induktion aus dem für $n = 2$. Für die Verteilungsfunktion F_X gilt:

$$\begin{aligned} 1 - F_X(t) = \Pr[X > t] &= \Pr[\min\{X_1, X_2\} > t] \\ &= \Pr[X_1 > t, X_2 > t] \\ &= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \\ &= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}. \end{aligned}$$

$\square$



4.4 Minimum von Exponentialverteilungen

Warten auf mehrere Ereignisse

Satz 88

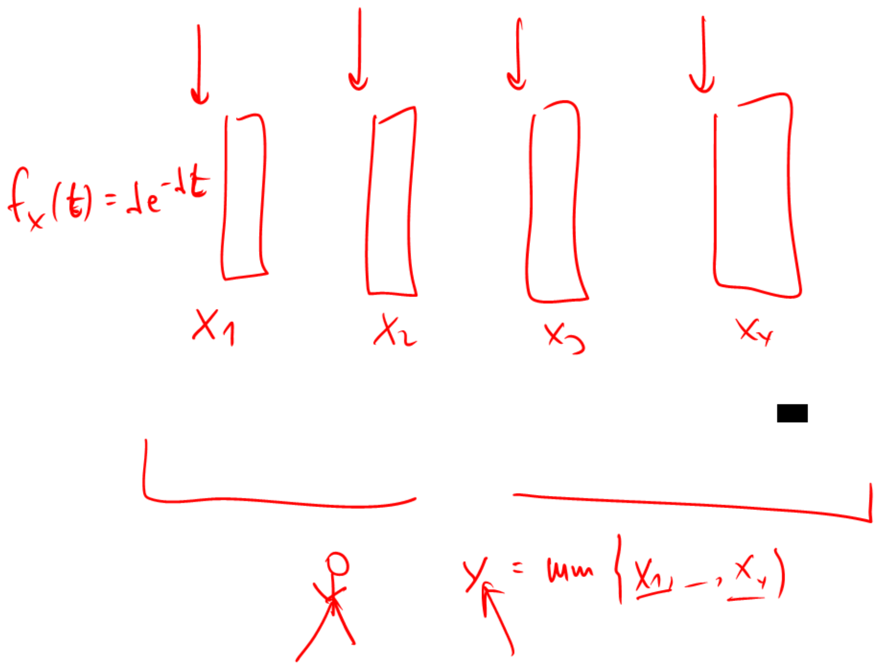
Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und exponentialverteilt mit den Parametern $\lambda_1, \ldots, \lambda_n$. Dann ist auch $X := \min\{X_1, \ldots, X_n\}$ exponentialverteilt mit dem Parameter $\lambda_1 + \ldots + \lambda_n$.

Beweis:

Der allgemeine Fall folgt mittels Induktion aus dem für $n = 2$. Für die Verteilungsfunktion F_X gilt:

$$\begin{aligned} 1 - F_X(t) = \Pr[X > t] &= \Pr[\min\{X_1, X_2\} > t] \\ &= \Pr[X_1 > t, X_2 > t] \\ &= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \\ &= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}. \end{aligned}$$

$\square$



4.4 Minimum von Exponentialverteilungen

Warten auf mehrere Ereignisse

Satz 88

Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und exponentialverteilt mit den Parametern $\lambda_1, \ldots, \lambda_n$. Dann ist auch $X := \min\{X_1, \ldots, X_n\}$ exponentialverteilt mit dem Parameter $\lambda_1 + \ldots + \lambda_n$.

Beweis:

Der allgemeine Fall folgt mittels Induktion aus dem für $n = 2$. Für die Verteilungsfunktion F_X gilt:

$$
\begin{aligned}
1 - F_X(t) = \Pr[X > t] &= \Pr[\min\{X_1, X_2\} > t] \\
&= \Pr[X_1 > t, X_2 > t] \\
&= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \\
&= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}.
\end{aligned}
$$

□

Anschaulich besagt Satz 88, dass sich die Raten addieren, wenn man auf das erste Eintreten eines Ereignisses aus mehreren unabhängigen Ereignissen wartet. Wenn beispielsweise ein Atom die Zerfallsrate λ besitzt, so erhalten wir bei n Atomen die Zerfallsrate $n\lambda$ (wie uns auch die Intuition sagt).

Poisson-Prozess

Wir hatten bei der Diskussion der geometrischen und der Poisson-Verteilung festgestellt:

Wenn der zeitliche Abstand der Treffer geometrisch verteilt ist, so ist ihre Anzahl in einer festen Zeitspanne binomialverteilt.

Im Grenzwert $n \to \infty$, wobei wir die Trefferwahrscheinlichkeit mit $p_n = \lambda/n$ ansetzen, konvergiert die Binomialverteilung gegen die Poisson-Verteilung und die geometrische Verteilung gegen die Exponentialverteilung. Im Grenzwert $n \to \infty$ erwarten wir deshalb die folgende Aussage:

Wenn man Ereignisse zählt, deren zeitlicher Abstand exponentialverteilt ist, so ist die Anzahl dieser Ereignisse in einer festen Zeitspanne Poisson-verteilt.

Seien $T_1, T_2 \ldots$ unabhängige exponentialverteilte Zufallsvariablen mit Parameter λ. Die Zufallsvariable T_i modelliert die Zeit, die zwischen Treffer $i-1$ und i vergeht.

Für den Zeitpunkt $t > 0$ definieren wir

$$X(t) := \max\{n \in \mathbb{N} \mid T_1 + \ldots + T_n \leq t\}.$$

$X(t)$ gibt also an, wie viele Treffer sich bis zur Zeit t (von Zeit Null ab) ereignet haben. Es gilt:

Seien $T_1, T_2 \ldots$ unabhängige exponentialverteilte Zufallsvariablen mit Parameter λ. Die Zufallsvariable T_i modelliert die Zeit, die zwischen Treffer $i-1$ und i vergeht.

Für den Zeitpunkt $t > 0$ definieren wir

$$X(t) := \max\{n \in \mathbb{N} \mid T_1 + \ldots + T_n \leq t\}.$$

$X(t)$ gibt also an, wie viele Treffer sich bis zur Zeit t (von Zeit Null ab) ereignet haben. Es gilt:

Fakt 89

Seien $T_1, T_2, \ldots$ unabhängige Zufallsvariablen und sei $X(t)$ für $t > 0$ wie oben definiert. Dann gilt: $X(t)$ ist genau dann Poisson-verteilt mit Parameter $t\lambda$, wenn es sich bei $T_1, T_2, \ldots$ um exponentialverteilte Zufallsvariablen mit Parameter λ handelt.

Zum Zufallsexperiment, das durch $T_1, T_2, \ldots$ definiert ist, erhalten wir für jeden Wert $t > 0$ eine Zufallsvariable $X(t)$. Hierbei können wir t als Zeit interpretieren und $X(t)$ als Verhalten des Experiments zur Zeit t. Eine solche Familie $(X(t))_{t>0}$ von Zufallsvariablen nennt man allgemein einen stochastischen Prozess. Der hier betrachtete Prozess, bei dem $T_1, T_2, \ldots$ unabhängige, exponentialverteilte Zufallsvariablen sind, heißt Poisson-Prozess und stellt ein fundamentales und zugleich praktisch sehr bedeutsames Beispiel für einen stochastischen Prozess dar.

Beispiel 90

Wir betrachten eine Menge von Jobs, die auf einem Prozessor sequentiell abgearbeitet werden. Die Laufzeiten der Jobs seien unabhängig und exponentialverteilt mit Parameter $\lambda = 1/30[1/s]$. Jeder Job benötigt also im Mittel $30s$.

Gemäß Fakt 89 ist die Anzahl von Jobs, die in einer Minute vollständig ausgeführt werden, Poisson-verteilt mit Parameter $t\lambda = 60 \cdot (1/30) = 2$.

Die Wahrscheinlichkeit, dass in einer Minute höchstens ein Job abgearbeitet wird, beträgt in diesem Fall ($t\lambda = 2$)

$$e^{-t\lambda} + t\lambda e^{-t\lambda} \approx 0{,}406 \,.$$

5. Verschiedene Approximationen der Binomialverteilung

Sei $H_n \sim \mathrm{Bin}(n,p)$ eine binomialverteilte Zufallsvariable mit der Verteilungsfunktion F_n. Für $n \to \infty$ gilt

$$\begin{aligned} F_n(t) &= \Pr[H_n/n \le t/n] \\ &\to \Phi\left(\frac{t/n - p}{\sqrt{p(1-p)/n}}\right) = \Phi\left(\frac{t - np}{\sqrt{p(1-p)n}}\right) . \end{aligned}$$

Wir können F_n somit für große n durch Φ approximieren. Diese Approximation ist in der Praxis deshalb von Bedeutung, da die Auswertung der Verteilungsfunktion der Binomialverteilung für große n sehr aufwendig ist, während für die Berechnung der Normalverteilung effiziente numerische Methoden vorliegen.

Beispiel 91

Wenn man die Wahrscheinlichkeit berechnen möchte, mit der bei 10^6 Würfen mit einem idealen Würfel mehr als 500500-mal eine gerade Augenzahl fällt, so muss man eigentlich folgenden Term auswerten:

$$T := \sum_{i=5{,}005\cdot 10^5}^{10^6} \binom{10^6}{i} \left(\frac{1}{2}\right)^{10^6} .$$

Dies ist numerisch kaum effizient möglich.

Die numerische Integration der Dichte φ der Normalverteilung ist hingegen relativ einfach. Auch andere Approximationen der Verteilung Φ, beispielsweise durch Polynome, sind bekannt. Entsprechende Funktionen werden in zahlreichen Softwarebibliotheken als „black box“ angeboten.

Beispiel

Mit der Approximation durch die Normalverteilung erhalten wir

$$\begin{aligned} T &\approx 1 - \Phi\left(\frac{5{,}005 \cdot 10^5 - 5 \cdot 10^5}{\sqrt{2{,}5 \cdot 10^5}}\right) \\ &= 1 - \Phi\left(\frac{5 \cdot 10^2}{5 \cdot 10^2}\right) \\ &= 1 - \Phi(1) \approx 0{,}1573 \,. \end{aligned}$$

Beispiel

Mit der Approximation durch die Normalverteilung erhalten wir

$$T \approx 1 - \Phi\left(\frac{5{,}005 \cdot 10^5 - 5 \cdot 10^5}{\sqrt{2{,}5 \cdot 10^5}}\right)$$
$$= 1 - \Phi\left(\frac{5 \cdot 10^2}{5 \cdot 10^2}\right)$$
$$= 1 - \Phi(1) \approx 0{,}1573\,.$$

Bei der Approximation der Binomialverteilung mit Hilfe von Korollar 82 führt man oft noch eine so genannte Stetigkeitskorrektur durch. Zur Berechnung von $\Pr[X \le x]$ für $X \sim \mathrm{Bin}(n,p)$ setzt man

$$\Pr[X \le x] \approx \Phi\left(\frac{x + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

statt

$$\Pr[X \le x] \approx \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right)$$

an.

$n = 10^6$ $\quad p = 1/2$

$t = 500.500$

Beispiel

Mit der Approximation durch die Normalverteilung erhalten wir

$$T \approx 1 - \Phi\left(\frac{5{,}005 \cdot 10^5 - 5 \cdot 10^5}{\sqrt{2{,}5 \cdot 10^4}}\right)$$
$$= 1 - \Phi\left(\frac{5 \cdot 10^2}{5 \cdot 10^2}\right)$$
$$= 1 - \Phi(1) \approx 0{,}1573\,.$$

Beispiel 76

25.05.2012, ZDF: Die Umfragen zu diesem Politbarometer wurden wie immer von der Mannheimer Forschungsgruppe Wahlen durchgeführt. Die Interviews wurden in der Zeit vom 22. bis 24. Mai 2012 bei 1312 zufällig ausgewählten Wahlberechtigten telefonisch erhoben. Die Befragung ist repräsentativ für die wahlberechtigte Bevölkerung in Deutschland. Der Fehlerbereich beträgt bei einem Parteianteil von 40 Prozent rund +/- drei Prozentpunkte und bei einem Parteianteil von zehn Prozent rund +/- zwei Prozentpunkte. Daten zur politischen Stimmung: CDU/CSU: 38 Prozent, SPD: 34 Prozent, FDP: zwei Prozent, Linke: vier Prozent, Grüne: 14 Prozent, Piraten: sechs Prozent.

Wie werden diese Fehlerbereiche berechnet?

- Approximation durch die Normalverteilung: Als Faustregel sagt man, dass die Verteilungsfunktion $F_n(t)$ von $\mathrm{Bin}(n,p)$ durch

$$F_n(t) \approx \Phi((t - np)/\sqrt{p(1-p)n})$$

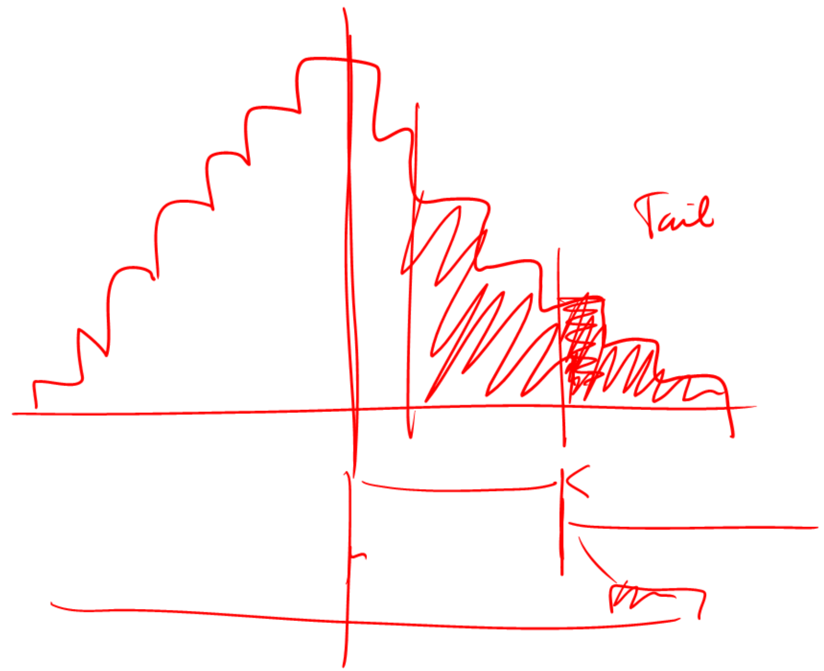
approximiert werden kann, wenn $np \geq 5$ und $n(1-p) \geq 5$ gilt.

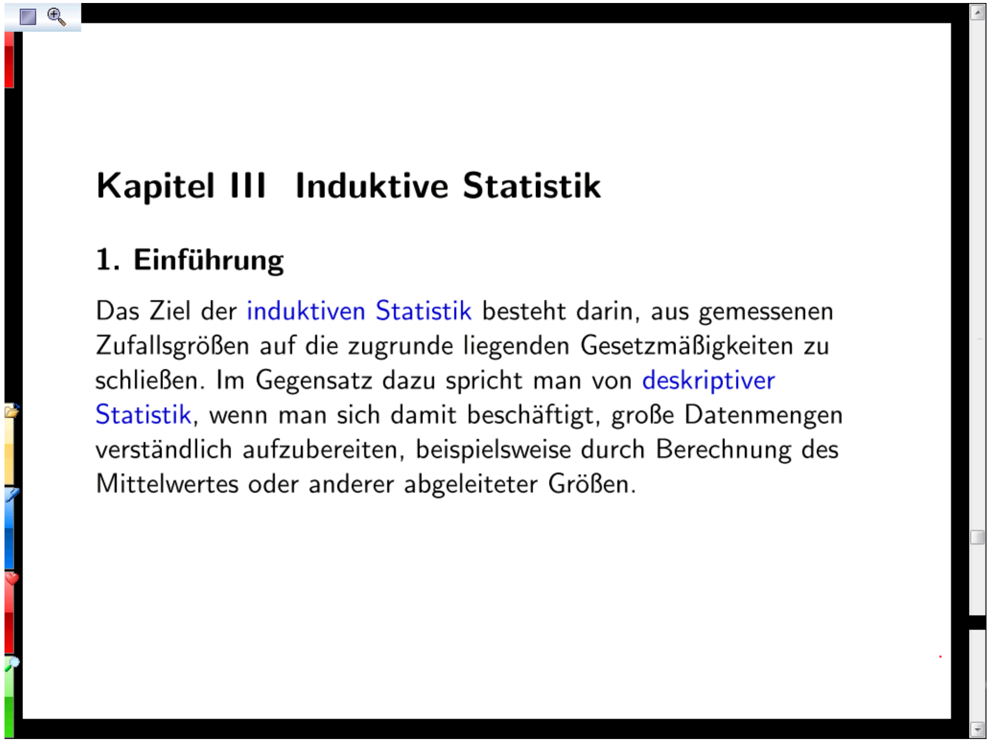
$n \to \infty$

$p \to 0$ $np = \lambda$

Approximationen für die Binomialverteilung

- Approximation durch die Poisson-Verteilung: $\mathrm{Bin}(n,p)$ wird approximiert durch $\mathrm{Po}(np)$. Diese Approximation funktioniert sehr gut für seltene Ereignisse, d. h. wenn np sehr klein gegenüber n ist. Als Faustregel fordert man $n \geq 30$ und $p \leq 0{,}05$.
- Approximation durch die Chernoff-Schranken: Bei der Berechnung der tails der Binomialverteilung liefern diese Ungleichungen meist sehr gute Ergebnisse. Ihre Stärke liegt darin, dass es sich bei den Schranken nicht um Approximationen, sondern um echte Abschätzungen handelt. Dies ist vor allem dann wichtig, wenn man nicht nur numerische Näherungen erhalten möchte, sondern allgemeine Aussagen über die Wahrscheinlichkeit von Ereignissen beweisen möchte.





Kapitel III Induktive Statistik

1. Einführung

Das Ziel der induktiven Statistik besteht darin, aus gemessenen Zufallsgrößen auf die zugrunde liegenden Gesetzmäßigkeiten zu schließen. Im Gegensatz dazu spricht man von deskriptiver Statistik, wenn man sich damit beschäftigt, große Datenmengen verständlich aufzubereiten, beispielsweise durch Berechnung des Mittelwertes oder anderer abgeleiteter Größen.