WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Seminar

Große Abweichungen
(WiSe 2010/2011)

Satz von Cramér für i.i.d. Multinomialverteilte Zufallsvariablen

Institut für Mathematische Statistik
Fachbereich Mathematik und Informatik
Westfälische Wilhelms-Universität Münster

vorgelegt von
Marius Staggenborg
(Matrikelnummer 339110)

Motivation: Die Theorie der Großen Abweichungen (Large Deviations) beschreibt, sehr allgemein gesagt, die Konvergenzgeschwindigkeit im schwachen Gesetz der großen Zahlen. Für Abweichungen mittlerer Größe gibt der Zentrale Grenzwertsatz eine erschöpfende Auskunft. Die Wahrscheinlichkeiten für größere Abweichungen gehen dagegen schnell nach Null und hier ist eben die genaue (exponentielle) Konvergenzgeschwindigkeit von Interesse.

Beispiel: Betrachten wir die Abschätzung einer $b(\cdot; n.p)$-verteilten Zufallsvariable S_n durch die Tchebyscheff-Ungleichung *(TSU)*, so erhalten wir eine Schranke der Form

$$P\left(|X - E(X)| \geq \epsilon\right) = P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{V\left(\frac{S_n}{n}\right)}{\epsilon^2} = \frac{p(1-p)}{n \cdot \epsilon^2}$$

Für das Beispiel eines symmetrischen Münzwurfes ($p = \frac{1}{2}$) mit $\epsilon = \frac{1}{10}$ und $n = 1000$

$$P\left(\left|\frac{S_{1000}}{n} - \frac{1}{2}\right| \geq \frac{1}{10}\right) \leq \frac{1}{40}$$

Wendet man statt der Tchebyscheff-Ungleichung eine alternative Form der Markoff-Ungleichung *(MKU)* an, sieht man, dass diese Abschätzung offensichtlich weit über der richtigen Wahrscheinlichkeit liegt. Benutzt man $\phi : x \to e^{\lambda x}$ mit $x \in \mathbb{R}^+, \lambda > 0$ beliebig, so erhalten wir

$$P\left(\left|\frac{S_n}{n}\right| \geq \alpha\right) = P\left(S_n \geq \alpha n\right) \leq \frac{E\left(\phi\left(|S_n|\right)\right)}{\phi\left(n\alpha\right)} = e^{-n\alpha\lambda} E\left(e^{\lambda S_n}\right)$$

Der Erwartungswert existiert, da S_n nur endlich viele Werte annehmen kann; diesen wollen wir nun auswerten. Es seien $X_1, \ldots, X_n$ unabhängige Zufallsgrößen mit $P\left(X_i = 1\right) = p$ und $P\left(X_i = 0\right) = (1-p)$. Wir schreiben dazu $\lambda S_n = \sum_{i=1}^{n} \lambda X_i$. Aufgrund der Unabhängigkeit der X_i folgt aus *Satz (1)* auch die Unabhänigkeit der $e^{\lambda X_i}$ und mit *Bemerkung (2)* für jedes $\lambda > 0$

$$P\left(S_n \geq \alpha n\right) \overset{(1),(2)}{\leq} \frac{E\left(\prod_{i=1}^{n} e^{\lambda X_i}\right)}{e^{n\alpha\lambda}} \overset{(1),(2)}{\leq} \frac{\prod_{i=1}^{n} E\left(e^{\lambda X_i}\right)}{e^{n\alpha\lambda}} = e^{-n\alpha\lambda}\left(E\left(e^{\lambda X_1}\right)\right)^n$$

Die Gleichheit gilt, da die Erwartungswerte der einzelnen X_i gleich sind, d.h. $E(X_1) = \ldots = E(X_n)$. Dies berechnen wir zu

$$P\left(S_n \geq \alpha n\right) \overset{(3)}{\leq} e^{-\alpha n \lambda}\left(pe^{\lambda}+(1-p)\right)^n = e^{n\left(-\alpha\lambda + log\left(pe^{\lambda}+(1-p)\right)\right)}$$

(3): $E\left(e^{\lambda X_i}\right) = e^{\lambda \cdot 1} P\left(X_i = 1\right) + e^{\lambda \cdot 0} P\left(X_i = 0\right) = pe^{\lambda} + (1-p)$
Sei nun $f(\lambda) = -\alpha\lambda + log\left(pe^{\lambda}+(1-p)\right)$. Wir wollen $\lambda > 0$ nun so wählen, dass wir eine möglichst gute obere Abschätzung erhalten. Dafür bestimmen wir das Minimum von f.

$$f'(\lambda) = -\alpha + \frac{pe^{\lambda}}{pe^{\lambda}+(1-p)} \qquad\qquad f''(\lambda) = \frac{p(1-p)e^{\lambda}}{\left(pe^{\lambda}+(1-p)\right)^2}$$

Wir sehen, dass für alle $\lambda > 0$ und $0 < p < 1$, $f''(\lambda) > 0$ gilt. $f'(\lambda)$ ist also streng monoton steigend.
$\Rightarrow$ Es existiert höchstens eine Nullstelle λ_0 von f', und f nimmt in diesem Punkt ihr absolutes Minimum an. Ist $\alpha \in (p, 1)$ - *(für $\alpha \geq 1$ ist $P\left(S_n \geq \alpha n\right) = 0$, und für $\alpha \leq p$ ist $\lambda_0 \leq 0$ was nicht sein kann wegen $\lambda > 0$)*, so ergibt sich aus $f'(\lambda_0) = 0$

$$0 = -\alpha + \frac{pe^{\lambda_0}}{pe^{\lambda_0}+(1-p)} = -\alpha pe^{\lambda_0} - \alpha(1-p) + pe^{\lambda_0} = pe^{\lambda_0}(1-\alpha) - \alpha(1-p)$$

$$\Leftrightarrow \qquad e^{\lambda_0} = \frac{\alpha(1-p)}{p(1-\alpha)} \qquad \Leftrightarrow \qquad \lambda_0 = \log\left(\frac{\alpha(1-p)}{p(1-\alpha)}\right) > 0$$

Setzen wir nun λ_0 in f ein erhalten wir

$$\begin{aligned}
f(\lambda_0) &= -\alpha \log\left(\frac{\alpha(1-p)}{p(1-\alpha)}\right) + \log\left(pe^{\log\left(\frac{\alpha(1-p)}{p(1-\alpha)}\right)} + (1-p)\right) \\
&= -\alpha\log\left(\frac{\alpha}{p}\right) + \alpha\log\left(\frac{1-\alpha}{1-p}\right) + \log\left(p\frac{\alpha(1-p)}{p(1-\alpha)} + (1-p)\right) \\
&= -\alpha\log\left(\frac{\alpha}{p}\right) + \alpha\log\left(\frac{1-\alpha}{1-p}\right) + \log\left(\frac{\alpha(1-p)}{1-\alpha} + \frac{(1-p)(1-\alpha)}{1-\alpha}\right) \\
&= -\alpha\log\left(\frac{\alpha}{p}\right) + \alpha\log\left(\frac{1-\alpha}{1-p}\right) + \log\frac{1-p}{1-\alpha} \\
&= -\alpha\log\left(\frac{\alpha}{p}\right) + \alpha\log\left(\frac{1-\alpha}{1-p}\right) - \log\frac{1-\alpha}{1-p} \\
&= -\alpha\log\left(\frac{\alpha}{p}\right) - (1-\alpha)\log\left(\frac{1-\alpha}{1-p}\right) =: -H\left(\alpha\,|p\right)
\end{aligned}$$

$H(\alpha\,|p)$ ist die *relative Entropie* von α bezüglich p und hat folgende Eigenschaften:

Lemma (4): Für $0 < p < 1$ ist $H(\cdot\,|p) \geq 0$ und $H(\alpha\,|p) = 0$ genau dann, wenn $\alpha = p$. Für ein Intervall $I = (a,b)$ ist $\inf_{\alpha \in I} H(\alpha\,|p) = 0$, falls $p \in I$. $H(\cdot\,|p)$ ist stetig und strikt konvex.

Beweis: Wir betrachten die folgende Hilfsfunktion $\psi(t) := t \log t - t + 1$ für $t > 0$ und $\psi(0) := 1$. Dann gilt: ψ ist nicht negativ, strikt konvex und $\psi(t) = 0 \Leftrightarrow t = 1$. Weiter gilt

$$H(\alpha|p) = p\psi\left(\frac{\alpha}{p}\right) + (1-p)\psi\left(\frac{1-\alpha}{1-p}\right)$$

Somit folgen die Eigenschaften jeweils aus den Eigenschaften der Funktion ψ. Wir betrachten exemplarisch den Beweis der Konvexität: Seien $\alpha_1, \alpha_2 \in (0,1)$ und $0 \leq \mu \leq 1$. Dann gilt mittels der Konvexität von ψ

$$\begin{aligned} H(\mu\alpha_1 + (1-\mu)\alpha_2|p) \leq p\mu\psi\left(\frac{\alpha_1}{p}\right) + p(1-\mu)\psi\left(\frac{\alpha_2}{p}\right) \\ &+ (1-p)\mu\psi\left(\frac{1-\alpha_1}{1-p}\right) + (1-p)(1-\mu)\psi\left(\frac{1-\alpha_2}{1-p}\right) \\ &= \mu H(\alpha_1|p) + (1-\mu)H(\alpha_2|p) \end{aligned}$$

□

Wir haben nun also gezeigt, dass für die Anzahl der „Kopf“-Würfe S_n, für alle $\alpha \in (p,1)$ die folgende Abschätzung gilt:

$$P\left(\frac{S_n}{n} \geq \alpha\right) \leq e^{-nH(\alpha|p)}$$

Für den symmetrischen Münzwurf gilt für alle $\alpha \in \left(0, \frac{1}{2}\right)$

$$P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \alpha\right) = 2P\left(\frac{S_n}{n} \geq \alpha + \frac{1}{2}\right) \leq 2e^{-nH\left(\alpha+\frac{1}{2}\middle|\frac{1}{2}\right)}$$

Für $\alpha = \frac{1}{10}$ und $n = 1000$ erhalten wir zum Beispiel

$$P\left(\left|\frac{S_{1000}}{1000} - \frac{1}{2}\right| \geq \frac{1}{10}\right) \leq 2e^{-nH\left(\alpha+\frac{1}{2}\middle|\frac{1}{2}\right)}$$

$$= 2e^{n\left(-\left(\alpha+\frac{1}{2}\right)\log\left(\frac{\alpha+\frac{1}{2}}{p}\right) - \left(1-\left(\alpha+\frac{1}{2}\right)\right)\log\left(\frac{1-\left(\alpha+\frac{1}{2}\right)}{1-p}\right)\right)}$$

$$= 2e^{1000\left(-\frac{6}{10}\log\left(\frac{12}{10}\right)-\frac{4}{10}\log\left(\frac{8}{10}\right)\right)} = 2e^{-600\log\left(\frac{6}{5}\right)}e^{-400\log\left(\frac{4}{5}\right)} = 2e^{600\log\left(\frac{5}{6}\right)}e^{400\log\left(\frac{5}{4}\right)}$$

$$= 2\left(\frac{5}{6}\right)^{600}\left(\frac{5}{4}\right)^{400} \leq 3,6 \cdot 10^{-9}$$

Diese Abschätzung ist offensichtlich viel besser als $\frac{1}{40}$ aus der Tchebyscheff-Ungleichung. Als nächstes wollen wir den Satz von Cramér für i.i.d. Multinomialverteilte Zufallsvariablen beweisen.

Satz und Definition (5): Zunächst definieren wir uns einige Funktionen. Sei eine Entropie-Funktion analog zum Fall r=2 definiert:

$$H(\rho|\pi) := \sum_{i=1}^{r} \rho_i \log\left(\frac{\rho_i}{\pi_i}\right) \text{ mit } \rho, \pi \in M(X) \text{ und } \pi_i > 0 \ \forall i$$

$M(X)$ ist hierbei wie folgt definiert,
$M(X) := \{(\rho_1, \ldots, \rho_r)|\rho_i \geq 0, \sum_{i=1}^{r} \rho_i = 1\} \subset \mathbb{R}^r$. Weiter sei $L_n(\omega, \cdot)$ der Vektor der relativen Häufigkeiten zu einer festen Beobachtung ω mit $L_n(\omega, i) = \frac{k_i}{n}$ für $i = 1, \ldots, r$. Wir wollen die Konvergenzgeschwindigkeit für unwahrscheinliche Ereignisse bestimmen, deshalb betrachten wir $P_\pi(L_n(\omega, \cdot) \in A)$ wobei P_π das n-fache Produktmaß ist und $A \subset M(X)$ die Menge der untypischen Ereignisse beschreibt. Formal

$$A := \{\nu \in M(X)| \ ||\nu - \pi||_{sup} \geq \epsilon\} \text{ mit } \epsilon > 0$$

dann gilt:

$$\lim_{n\to\infty} \frac{1}{n} \log\left(P_\pi\left(L_n(\omega, \cdot) \in A\right)\right) = -\inf_{\rho\in A}\left(H(\rho|\pi)\right)$$

Beweis: Wegen der Multinomialverteilung von L_n gilt:

$$P_\pi(L_n(\omega, \cdot) \in A) = \sum_{k=(k_1,\ldots,k_r)\in E_n} \frac{n!}{k_1! \cdots k_r!} \pi_1^{k_1} \cdots \pi_r^{k_r}$$

Wobei E_n die Menge aller möglichen, zulässigen - im Sinne der Definition von A (unwahrscheinliche) - Konfigurationen nach n Ziehungen beschreibt.

$$E_n = \{(k_1, \ldots, k_r)| \sum_{i=1}^{r} k_i = n, k_i \in \{0, \ldots, n\} \text{ und } ||\frac{k}{n} - \pi||_{sup} \geq \epsilon\}$$

Wie im Fall der Binomialverteilung erhalten wir folgende Abschätzung:

$$\max_{k\in E_n}\left(\frac{n!}{k_1!\cdots k_r!}\pi_1^{k_1}\cdots\pi_r^{k_r}\right)\le P_\pi(L_n(\omega,\cdot)\in A)$$

$$=\sum_{k=(k_1,\dots,k_r)\in E_n}\frac{n!}{k_1!\cdots k_r!}\pi_1^{k_1}\cdots\pi_r^{k_r}\le(n+1)^r\max_{k\in E_n}\left(\frac{n!}{k_1!\cdots k_r!}\pi_1^{k_1}\cdots\pi_r^{k_r}\right)$$

Erweitern wir nun mit $\frac{1}{n}$ und wenden den Logarithmus an erhalten wir

$$\max_{k\in E_n}\frac{1}{n}\log\left(\frac{n!}{k_1!\cdots k_r!}\pi_1^{k_1}\cdots\pi_r^{k_r}\right)\le\frac{1}{n}\log\left(P_\pi(L_n(\omega,\cdot)\in A)\right)$$

$$\le\frac{1}{n}\log\left((n+1)^r\max_{k\in E_n}\left(\frac{n!}{k_1!\cdots k_r!}\pi_1^{k_1}\cdots\pi_r^{k_r}\right)\right)$$

$$=\frac{1}{n}\log(n+1)^r+\max_{k\in E_n}\frac{1}{n}\log\left(\frac{n!}{k_1!\cdots k_r!}\pi_1^{k_1}\cdots\pi_r^{k_r}\right)$$

Das Maximum können wir rausziehen, da der Logarithmus monoton ist. Wir versuchen uns nun langsam an obigen Term heranzutasten. Mit der Stirling-Formel $\ln(n!)=n\ln(n)-n+O(\log(n))$ folgt zunächst:

$$\frac{1}{n}\log\left(\frac{n!}{k_1!\cdots k_r!}\right)=\frac{1}{n}\left(\log(n!)-\log(k_1!\cdots k_r!)\right)$$

$$\overset{Stirling}{=}\frac{1}{n}\left(n\log(n)-n+O(\log(n))-\log(k_1!\cdots k_r!)\right)$$

$$=\log(n)-1+\frac{O(\log(n))}{n}-\frac{1}{n}\log(k_1!\cdots k_r!)$$

$$=\log(n)-1+O\left(\frac{\log(n)}{n}\right)-\frac{1}{n}\sum_{j=1}^{r}\log(k_j!)$$

$$\overset{Stirling}{=}\log(n)-1+O\left(\frac{\log(n)}{n}\right)-\frac{1}{n}\left(\sum_{j=1}^{r}k_j\log(k_j)-\sum_{j=1}^{r}k_j+\sum_{j=1}^{r}O(\log(k_j))\right)$$

$$=\log(n)-\sum_{j=1}^{r}\frac{k_j}{n}\log(k_j)+O\left(\frac{\log(n)}{n}\right)$$

$$\overset{(6)}{=}-\sum_{j=1}^{r}\frac{k_j}{n}\log\left(\frac{1}{n}\right)-\sum_{j=1}^{r}\frac{k_j}{n}\log(k_j)+O\left(\frac{\log(n)}{n}\right)$$

$$=-\sum_{j=1}^{r}\left(\frac{k_j}{n}\log\left(\frac{1}{n}\right)+\frac{k_j}{n}\log(k_j)\right)+O\left(\frac{\log(n)}{n}\right)$$

$$= -\sum_{j=1}^{r} \frac{k_j}{n} \log\left(\frac{k_j}{n}\right) + O\left(\frac{\log(n)}{n}\right)$$

(6): $-\sum\limits_{j=1}^{r} \frac{k_j}{n} \log\left(\frac{1}{n}\right) = -\sum\limits_{j=1}^{r} \log\left(\left(\frac{1}{n}\right)^{\frac{k_j}{n}}\right) = -\log\left(\frac{1}{n}^{\frac{k_1+\ldots+k_r}{n}}\right) = \log(n),$

da $k_1 + \ldots + k_r = n$. Wir erweitern unseren Term nun um $\pi_1^{k_1} \cdots \pi_r^{k_r}$. Dann ergibt sich

$$\frac{1}{n} \log\left(\frac{n!}{k_1! \cdots k_r!} \pi_1^{k_1} \cdots \pi_r^{k_r}\right) = \frac{1}{n} \log\left(\frac{n!}{k_1! \cdots k_r!}\right) + \frac{1}{n} \log\left(\pi_1^{k_1} \cdots \pi_r^{k_r}\right)$$

$$= -\sum_{j=1}^{r} \frac{k_j}{n} \log\left(\frac{k_j}{n}\right) + \sum_{j=1}^{r} \frac{1}{n} \log(\pi_j^{k_j}) + O\left(\frac{\log(n)}{n}\right)$$

$$= -\sum_{j=1}^{r} \frac{k_j}{n} \log\left(\frac{k_j}{n}\right) + \sum_{j=1}^{r} \frac{k_j}{n} \log(\pi_j) + O\left(\frac{\log(n)}{n}\right)$$

$$= \sum_{j=1}^{r} \frac{k_j}{n} \left(\log(\pi_j) - \log\left(\frac{k_j}{n}\right)\right) + O\left(\frac{\log(n)}{n}\right)$$

$$= -H\left(\rho_{\frac{k}{n}} | \pi\right) + O\left(\frac{\log(n)}{n}\right)$$

$\rho_{\frac{k}{n}}$ ist hierbei der Vektor mit den Einträgen $\frac{k_i}{n}$. Dies können wir nun in unsere erste Ungleichung einsetzen und weiter abschätzen.

$$\max_{k \in E_n} -H\left(\rho_{\frac{k}{n}} | \pi\right) - O\left(\frac{\log(n)}{n}\right) \leq \max_{k \in E_n} \frac{1}{n} \log\left(\frac{n!}{k_1! \cdots k_r!} \pi_1^{k_1} \cdots \pi_r^{k_r}\right)$$

$$\leq \max_{k \in E_n} -H\left(\rho_{\frac{k}{n}} | \pi\right) + O\left(\frac{\log(n)}{n}\right)$$

Die Menge $\{\rho \in M(X) | \rho = \rho_{\frac{k}{n}}, k \in E_n\}$ ist kompakt und liegt für $n \to \infty$ dicht in A. Da $H(\cdot|\pi)$ stetig ist und jede stetige Funktion auf kompakten Mengen ihr Maximum annimmt, folgt weiter:

$$\max_{k \in E_n} -H\left(\rho_{\frac{k}{n}} | \pi\right) \to_{n \to \infty} \max_{\rho \in A} -H\left(\rho | \pi\right)$$

Insgesamt erhalten wir folgende Gleichung:

$$\lim_{n \to \infty} \frac{1}{n} \log\left(P_\pi\left(L_n(\omega, \cdot) \in A\right)\right) = \max_{\rho \in A}\left(-H(\rho|\pi)\right) = -\inf_{\rho \in A}\left(H(\rho|\pi)\right)$$

$\square$

Satz (1): Sind die Zufallsgrößen $X_1, \ldots, X_n$ unabhängig und sind $f_i : \mathbb{R} \to \mathbb{R}$ für $i = 1, \ldots, n$ beliebige Funktionen, so sind die Zufallsgrößen $Y_i = f_i \circ X_i$ für $i = 1, \ldots, n$ unabhängig.

Bemerkung (2): Für n unabhänige Zufallsgrößen $X_1, \ldots, X_n$ mit existierenden Erwartungswerten, existiert auch der Erwartungswert von $\prod_{i=1}^{n} X_i$ und es gilt $E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E\left(X_i\right)$.

MKU: Sei ϕ eine nicht negative, monoton wachsende Funktion auf $[0, \infty)$. Weiter sei X eine Zufallsgröße, für die der Erwartungswert $E\left(\phi\left(|X|\right)\right)$ existiert. Dann gilt für jedes $a > 0$ mit $\phi(a) > 0$

$$(P\left(|X| \geq \alpha\right) \leq \frac{E\left(\phi(|X|)\right)}{\phi(\alpha)})$$

TSU: Sei X eine Zufallsgröße, deren Erwartungswert $E(X)$ und Varianz $V(X)$ existieren. Dann gilt für jedes $a > 0$

$$P\left(|X - E(X)| \geq a\right) \leq \frac{V(X)}{a^2}$$