

1.2 Summen von Zufallsvariablen aus einer Zufallsstichprobe

Nachdem eine Stichprobe $X_1, \dots, X_n$ gezogen wurde berechnen wir gewöhnlich irgendwelchen Wert damit. Sei dies $T = T(X_1, \dots, X_n)$, wobei T auch vektorwertig sein kann.

Die Verteilung der Zufallsvariablen T wird in Termen der Populationsverteilung beschrieben aus der die Stichprobe stammt. Im Folgenden diskutieren wir Eigenschaften von Funktionen $T(X_1, \dots, X_n)$, die als Summen von Zufallsvariablen definiert sind.

Definition 1.2.1: Sei $X_1, \dots, X_n$ eine Zufallsstichprobe vom Umfang n und sei $T(x_1, \dots, x_n)$ eine skalar- oder vektorwertige Funktion deren Definitionsbereich den Stichprobenraum von $X_1, \dots, X_n$ beinhaltet. Die Zufallsvariable oder den Zufallsvektor $T = T(X_1, \dots, X_n)$ nennt man eine **Statistik** und deren Verteilung bezeichnet man als **Stichprobenverteilung** von T .

Diese Definition einer Statistik ist sehr breit mit der einzigen Einschränkung, dass T keine Funktion von Parametern sein darf. Jetzt definieren wir einige häufig verwendete Statistiken.

Definition 1.2.2: Das **Stichprobenmittel** (empirische Mittel, sample mean) ist definiert als

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

Definition 1.2.3: Die **Stichprobenvarianz** (empirische Varianz, sample variance) ist definiert als

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 .$$

Die **Stichprobenstandardabweichung** ist $S = \sqrt{S^2}$.

Die beobachteten Werte dieser Statistiken bezeichnet man durch $\bar{x}$, s^2 , und s .

Satz 1.2.1: Seien $x_1, \dots, x_n$ beliebige Zahlen und sei $\bar{x} = \sum_i x_i/n$. Dann gilt

(a) $\min_a \sum_i (x_i - a)^2 = \sum_i (x_i - \bar{x})^2,$

(b) $(n - 1)s^2 = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - n\bar{x}^2.$

Lemma 1.2.1: Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer Population, und sei $g(x)$ eine Funktion sodass $E(g(X_1))$ und $\text{var}(g(X_1))$ existieren. Dann gilt

$$\begin{aligned} E \left(\sum_{i=1}^n g(X_i) \right) &= nE(g(X_1)), \\ \text{var} \left(\sum_{i=1}^n g(X_i) \right) &= n\text{var}(g(X_1)). \end{aligned}$$

Satz 1.2.2: Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer Population mit Erwartung μ und Varianz $\sigma^2 < \infty$. Dann gilt

(a) $E(\bar{X}) = \mu$ (auch für $\sigma^2 = \infty$),

(b) $\text{var}(\bar{X}) = \sigma^2/n$,

(c) $E(S^2) = \sigma^2$.

Bemerkungen:

$\bar{X}$ und S^2 sind Beispiele von einer **unverzerrten (unbiased)** Statistik. $\bar{X}$ ist unverzerrt für μ und S^2 ist unverzerrt für σ^2 . Die Skalierung von S^2 mit $(n - 1)$ ist daher sinnvoll gewählt. Es gilt damit $E(S^2) = \sigma^2$.

Wäre S^2 nur mit n skaliert, so würde dafür folgen

$$E \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = E \left(\frac{1}{n} (n - 1) S^2 \right) = \frac{n - 1}{n} \sigma^2$$

und dies wäre nicht mehr unverzerrt für σ^2 (jedoch wäre dies noch immer asymptotisch unverzerrt, für $n \rightarrow \infty$).

Wir diskutieren nun die Stichprobenverteilung von $\bar{X}$.

Satz 1.2.3: Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer Population mit Momentenerzeugender Funktion $M_X(t)$. Dann ist die Momentenerzeugende Funktion des Stichprobenmittels gleich

$$M_{\bar{X}}(t) = (M_X(t/n))^n.$$

Beispiel 1.2.1: Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer $N(\mu, \sigma^2)$ Population. Dann ist $M_{X_i}(t) = \exp(\mu t + \sigma^2 t^2/2)$ für alle $i = 1, \dots, n$ und es folgt

$$\begin{aligned} M_{\bar{X}}(t) &= \left(\exp \left(\mu \frac{t}{n} + \frac{\sigma^2 (t/n)^2}{2} \right) \right)^n \\ &= \exp \left(n \left(\mu \frac{t}{n} + \frac{\sigma^2 (t/n)^2}{2} \right) \right) = \exp \left(\mu t + \frac{(\sigma^2/n) t^2}{2} \right), \end{aligned}$$

d.h., $\bar{X} \sim N(\mu, \sigma^2/n)$.

Falls wir eine Stichprobe aus einer Population ziehen, die ein Mitglied der *Exponentialfamilie* ist, so ist die Stichprobenverteilung von Summen von Zufallsvariablen, und speziell von $\overline{X}$, sehr einfach zu erhalten.

Definition 1.2.4: Eine Familie von Dichte- oder Wahrscheinlichkeitsfunktionen nennt man **Exponentialfamilie**, falls sie dargestellt werden kann als

$$f(x|\boldsymbol{\theta}) = h(x) c(\boldsymbol{\theta}) \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) t_i(x) \right) .$$

Hierbei sind $h(x) \geq 0$ und $t_1(x), \dots, t_k(x)$ reellwertige Funktionen der Beobachtung x (von $\boldsymbol{\theta}$ unabhängig), und $c(\boldsymbol{\theta}) \geq 0$ und $w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})$ sind reellwertige Funktionen des (möglicherweise vektorwertigen) Parameters $\boldsymbol{\theta}$ (von x unabhängig).

Viele bekannte Verteilungsfamilien gehören zur Exponentialfamilie: Normal, Gamma und Beta (stetig), sowie Binomial, Poisson und Negativ-Binomial (diskret).

Um zu prüfen, ob eine Familie $f(x|\boldsymbol{\theta})$ Mitglied der Exponentialfamilie ist, muss man die Funktionen $h(x)$, $c(\boldsymbol{\theta})$, $w_i(\boldsymbol{\theta})$, und $t_i(x)$ identifizieren und zeigen, dass diese Familie die Form aus Definition 1.2.4 hat.

Beispiel 1.2.2: Sei n ein positiver ganzzahliger Wert und betrachte die Binomial(n, p) Familie mit $0 < p < 1$. Dann ist die Wahrscheinlichkeitsfunktion für diese Familie für $x = 0, 1, \dots, n$ gleich

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{n}{x} (1-p)^n \left(\frac{p}{1-p} \right)^x \\ &= \binom{n}{x} (1-p)^n \exp \left(\log \left(\frac{p}{1-p} \right) \cdot x \right) . \end{aligned}$$

Definiere

$$h(x) = \begin{cases} \binom{n}{x} & x = 0, 1, \dots, n \\ 0 & \text{sonst} \end{cases}, \quad c(p) = (1-p)^n, \quad 0 < p < 1$$

$$w_1(p) = \log \frac{p}{1-p}, \quad 0 < p < 1, \quad t_1(x) = x.$$

Damit haben wir die äquivalente Darstellung

$$f(x|p) = h(x)c(p) \exp(w_1(p)t_1(x))$$

was eine Exponentialfamilie ist mit $k = 1$.

Bemerke, dass $h(x) > 0$ nur für $x = 0, 1, \dots, n$, und dass $c(p)$ nur für $0 < p < 1$ definiert ist. Dies ist wichtig, da die obige Form für alle Werte von x mit der ursprünglichen übereinstimmen muss.

Weiters ist $f(x|p)$ nur für $0 < p < 1$ eine Exponentialfamilie (weil nur hierfür die Funktionen in den Parametern definiert sind).

Manchmal sind auch die Werte $p = 0$ und $p = 1$ im Binomialmodell inkludiert; jedoch nicht hier, da der Bereich der x Werte für die $f(x|p) > 0$ gilt unter $p = 0/1$ von den entsprechenden Bereichen unter den übrigen p Werten unterscheidet.

Wir betrachten nun statistische Eigenschaften (Momente) der Exponentialfamilie.

Satz 1.2.4: Sei X eine Zufallsvariable mit Dichte- oder Wahrscheinlichkeitsfunktion aus der Exponentialfamilie. Dann gilt

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^k \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X) \right) &= -\frac{\partial}{\partial \theta_j} \log(c(\boldsymbol{\theta})), \\ \text{var} \left(\sum_{i=1}^k \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X) \right) &= -\frac{\partial^2}{\partial \theta_j^2} \log(c(\boldsymbol{\theta})) - \mathbb{E} \left(\sum_{i=1}^k \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(X) \right). \end{aligned}$$

Beispiel 1.2.2 Fortsetzung: Berechne mit Satz 1.2.4 die Momente der Binomial(n, p) Verteilung, $0 < p < 1$.

$$\frac{d}{dp} w_1(p) = \frac{d}{dp} \log \frac{p}{1-p} = \frac{1}{p(1-p)}$$

$$\frac{d}{dp} \log c(p) = \frac{d}{dp} \log (1-p)^n = -\frac{n}{1-p}$$

Damit resultiert

$$E \left(\frac{1}{p(1-p)} X \right) = \frac{n}{1-p}$$

also die bekannte Eigenschaft $E(X) = np$.

Ähnliches Vorgehen liefert die bekannte Varianzidentität $\text{var}(X) = np(1-p)$.

Beispiel 1.2.3: Sei $f(x|\mu, \sigma^2)$ die $\text{Normal}(\mu, \sigma^2)$ Familie von Dichten und betrachte $\theta = (\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma > 0$, also

$$\begin{aligned} f(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right). \end{aligned}$$

Definiere

$$\begin{aligned} h(x) &= 1, \quad x \in \mathbb{R}, \\ c(\theta) = c(\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right), \quad \mu \in \mathbb{R}, \sigma > 0; \\ w_1(\theta) &= \frac{1}{\sigma^2}, \quad \sigma > 0, \quad w_2(\theta) = \frac{\mu}{\sigma^2}, \quad \sigma > 0, \\ t_1(x) &= -x^2/2, \quad t_2(x) = x. \end{aligned}$$

Damit hat

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma^2) \exp (w_1(\mu, \sigma^2)t_1(x) + w_2(\mu, \sigma^2)t_2(x))$$

genau die ursprüngliche Form mit $k = 2$ und die Parameterfunktionen sind nur über den Bereich der Parameter definiert.

Im Allgemeinen darf in einer Exponentialfamilie der Bereich der x Werte, für den $f(x|\boldsymbol{\theta}) > 0$ gilt, nicht von $\boldsymbol{\theta}$ abhängen. Die gesamte Definition der Dichte oder Wahrscheinlichkeitsfunktion muss in die Form $f(x|\boldsymbol{\theta})$ aus Definition 1.2.4 eingeflossen sein. Am einfachsten kann der zulässige Bereich der x Werte mithilfe von Indikatoren charakterisiert werden, d.h. mittels

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} , \quad \text{oder} \quad I(x \in A) .$$

Für die Normalverteilungsfamilie folgt damit

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma^2) \exp(w_1(\mu, \sigma^2)t_1(x) + w_2(\mu, \sigma^2)t_2(x))I_{\mathbb{R}}(x)$$

Da hier der Indikator nur von x abhängt, kann man ihn in $h(x)$ inkludieren.

Bemerkungen:

Da der Faktor $\exp(\cdot)$ immer positiv ist, gilt für beliebiges $\theta \in \Theta$ (Θ bezeichnet hierbei den gesamten Parameterraum), also für ein beliebiges θ mit $c(\theta) > 0$, dass $\{x : f(x|\theta) > 0\} = \{x : h(x) > 0\}$ und diese Menge hängt nicht von θ ab.

So ist z.B. $f(x|\theta) = \theta^{-1} \exp(1 - x/\theta)$, $0 < \theta < x < \infty$, **keine** Exponentialfamilie obwohl wir (fälschlicherweise) schreiben könnten

$$f(x|\theta) = h(x)c(\theta) \exp(w(\theta)t(x))$$

mit $h(x) = e^1$, $c(\theta) = \theta^{-1}$, $w(\theta) = \theta^{-1}$ und $t(x) = -x$.

Mit dem darin fehlenden Indikator wird dies klar:

$$f(x|\theta) = \theta^{-1} \exp(1 - x/\theta) I_{(\theta, \infty)}(x) .$$

Diesen Indikator darf man weder in $h(\cdot)$ noch in $c(\cdot)$ geben, da er eine Funktion sowohl in θ als auch in x ist.

Definition 1.2.5: Eine **gekrümmte** (curved) Exponentialfamilie ist eine Familie von Dichten aus der Exponentialfamilie, für welche die Dimension des Vektors θ gleich $d < k$ ist. Ist $d = k$, nennt man die Familie eine **volle** (full) Exponentialfamilie.

Beispiel 1.2.4: Die $N(\mu, \sigma^2)$ Familie ist eine volle Familie ($k = 2$). Für beispielsweise $\sigma^2 = \mu^2$ wird diese gekrümmt. Dann hat man

$$\begin{aligned} f(x|\mu) &= \frac{1}{\sqrt{2\pi\mu^2}} \exp\left(-\frac{(x-\mu)^2}{2\mu^2}\right) \\ &= \frac{1}{\sqrt{2\pi\mu^2}} \exp\left(-\frac{1}{2}\right) \exp\left(-\frac{x^2}{2\mu^2} + \frac{x}{\mu}\right). \end{aligned}$$

Für die volle Exponentialfamilie haben wir hier als Parameterraum

$$(\mu, \sigma^2) = \mathbb{R} \times (0, \infty),$$

während für die obige gekrümmte Version dafür folgt

$$(\mu, \sigma^2) = (\mu, \mu^2),$$

eine Parabel.

Normal Approximationen ergeben häufig gekrümmte Exponentialfamilien:

Für $X_i \stackrel{iid}{\sim} \text{Bernoulli}(p)$ folgt approximativ (ZGWS)

$$\overline{X} \sim N \left(p, \frac{1}{n} p(1 - p) \right) .$$

Für $X_i \stackrel{iid}{\sim} \text{Poisson}(\lambda)$ ergibt sich approximativ (ZGWS)

$$\overline{X} \sim N \left(\lambda, \frac{1}{n} \lambda \right) .$$

Satz 1.2.5: Angenommen $X_1, \dots, X_n$ ist eine Zufallsstichprobe aus einer Population mit Dichte- oder Wahrscheinlichkeitsfunktion $f(x|\boldsymbol{\theta})$ der Form

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) t_i(x) \right) .$$

Definiere Statistiken $T_1, \dots, T_k$ durch

$$T_i(X_1, \dots, X_n) = \sum_{j=1}^n t_i(X_j), \quad i = 1, \dots, k .$$

Falls die Menge $\{(w_1(\boldsymbol{\theta}), \dots, w_k(\boldsymbol{\theta})), \boldsymbol{\theta} \in \Theta\}$ eine offene Teilmenge des $\mathbb{R}^k$ enthält, dann ist die Verteilung von $(T_1, \dots, T_k)$ eine Exponentialfamilie der Form

$$f_T(u_1, \dots, u_k|\boldsymbol{\theta}) = H(u_1, \dots, u_k) (c(\boldsymbol{\theta}))^n \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) u_i \right) .$$

Bemerkungen:

Die obige Bedingung (offene Teilmenge) eliminiert Dichten wie die der $N(\theta, \theta^2)$ -Verteilung, und eliminiert im allgemeinen gekrümmte Exponentialfamilien.

Auch hier sind die Funktionen $c(\theta)$ und $w_i(\theta)$ dieselben wie in der Originalfamilie. Die Funktion $H(\cdot)$ jedoch unterscheidet sich von $h(\cdot)$.

Beispiel 1.2.5: Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer Bernoulli(p) Verteilung. Aus dem Beispiel 1.2.2 sehen wir (mit Wahl $n = 1$), dass die Bernoulli(p) Verteilung eine Exponentialfamilie ist mit $k = 1$, $c(p) = (1 - p)$, $w(p) = \log p / (1 - p)$ und $t(x) = x$.

Somit ist in Satz 1.2.5 die Statistik $T_1 = T_1(X_1, \dots, X_n) = \sum_{j=1}^n X_j$.

Wir wissen, dass $T_1 \sim \text{Binomial}(n, p)$ gilt. Von Beispiel 1.2.2 wissen wir, dass die Binomial(n, p) eine Exponentialfamilie ist mit gleichem $w(p)$ und $c(p) = (1 - p)^n$. Somit ist zumindest die Aussage des Satzes 1.2.5 für diesen Fall verifiziert.