K 16 Verteilungen für die Statistik

Seien

$$X_1, \ldots, X_n$$

Zufallsvariablen und

$$W_1, \ldots, W_m$$

Statistiken, die aus $\{X_i\}$ berechnet sind. z.B.

$$W_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$W_2 = \max\{X_1, \ldots, X_n\}$$

$$W_3 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Jede W_j ist eine Transformation der $\{X_i\}$ und deshalb auch eine ZV. Wenn wir annehmen, dass die $\{X_i\}$ u.i.v. sind, können wir im Prinzip die Verteilungen von allen möglichen W_j ableiten.

Um Stichprobenwerte zu beurteilen, brauchen wir die Verteilungen der entsprechenden Statistiken.

Statistiken aus Stichproben

Sei X eine ZV mit Dichte f und Verteilungsfunktion F.

Seien $\{X_i; i = 1, \ldots n\}$ eine zufällige Stichprobe aus X.

$$\Rightarrow \{X_i\} \qquad \text{u.i.v.}$$

Statistiken sind Zusammenfassungen der Daten aus der Stichprobe.

16.2 Resultate für alle Verteilungen

16.2.1 Verteilung von $\bar{X}$

Der ZGS besagt, dass für $n \to \infty$

$$\bar{X} \to N(\mu, \frac{\sigma^2}{n})$$

unabhängig von der Form der Verteilung der $\{X_i\}$.

16.2.2 Verteilung von $\max X$

Für X_i mit Verteilungsfunktion $F(x)$ hat

$$Y = \max X$$

die Verteilungsfunktion

$$G(y) = F(y)^n$$

16.3 Die χ_k^2 Verteilung für X^2 und Quadratsummen

$$X \sim N(0,1) \quad \Rightarrow \quad Y = X^2 \sim \chi_1^2$$

Seien $g(y)$ die Dichte von $Y = X^2$ und $G(y)$ die Verteilungsfunktion von Y.

$$\begin{aligned} G_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= 1 - 2\Phi(-\sqrt{y}) \\ g(y) &= \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \end{aligned}$$

eine χ_1^2 Dichte.

Die Chiquadratverteilung mit k Freiheitsgraden, χ_k^2, hat die Dichte

$$f(x) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}}$$

mit Erwartungswert k und Varianz $2k$.

Die charakteristische Funktion für $Y \sim \chi_k^2$ ist

$$(1-2it)^{-\frac{k}{2}}$$

Für die Chiquadratverteilung gilt

$$\text{wenn} \quad Y_1 \sim \chi_{k_1}^2 \quad Y_2 \sim \chi_{k_2}^2 \quad \text{unabhängig}$$

$$Y_1 + Y_2 \sim \chi_{k_1+k_2}^2$$

weil aus den c.f.'s folgt

$$(1-2it)^{-\frac{k_1}{2}} * (1-2it)^{-\frac{k_2}{2}} = (1-2it)^{-\frac{k_1+k_2}{2}}$$

16.3 Die Gammaverteilung

Eine gammaverteilte ZV $\Gamma(\lambda, \nu)$ hat die Dichte

$$f(y) = \frac{1}{\Gamma(\nu)} \lambda^{\nu} y^{\nu-1} e^{-\lambda y}$$

$$\Rightarrow \chi_1^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \quad \text{weil} \quad \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

$$\chi_2^2 = \Gamma\left(\frac{1}{2}, 1\right) = E(\frac{1}{2})$$

$$\chi_k^2 = \Gamma\left(\frac{1}{2}, \frac{k}{2}\right)$$

Der Erwartungswert und Varianz von $Y \sim \Gamma(\lambda, \nu)$ sind

$$E[Y] = \frac{\nu}{\lambda} \quad \text{und} \quad V[Y] = \frac{\nu}{\lambda^2}$$

Die charakteristische Funktion für $Y \sim \Gamma(\lambda, \nu)$ ist

$$\left(1 - \frac{it}{\lambda}\right)^{-\nu}$$

Die Gammaverteilung hat auch die wichtige Eigenschaft, dass

$$\text{für} \quad Y_1 \sim \Gamma(\lambda, \nu_1) \qquad Y_2 \sim \Gamma(\lambda, \nu_2) \quad \text{unabhängig}$$

$$Y_1 + Y_2 \sim \Gamma(\lambda, \nu_1 + \nu_2)$$

weil aus den c.f.'s folgt

$$\left(1 - \frac{it}{\lambda}\right)^{-\nu_1} * \left(1 - \frac{it}{\lambda}\right)^{-\nu_2} = \left(1 - \frac{it}{\lambda}\right)^{-\nu_1 - \nu_2}$$

16.4 Die Verteilung von s^2 für $X \sim N(0,1)$

$$\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

wenn alle $\{X_i\}$ u.i.v. $\sim N(0,1)$, weil die Summe von zwei unabhängig χ^2-verteilte ZV auch χ^2-verteilt ist. Aber

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

ist NICHT die Summe von n unabhängig verteilten ZV.

Helmert hat die folgende orthogonale Transformation vorgeschlagen:

$$\begin{aligned}
U_1 &= \frac{1}{\sqrt{2}}(X_1 - X_2) \\
U_2 &= \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3) \\
&\vdots \\
U_{n-1} &= \frac{1}{\sqrt{n(n-1)}}(X_1 + \ldots + X_{n-1} - (n-1)X_n) \\
U_n &= \frac{1}{\sqrt{n}} \sum X_i
\end{aligned}$$

Orthogonal heißt für $U_i = \sum_{j=1}^{n} c_{ij} X_j$ gilt $\sum_{i=1}^{n} c_{ij} c_{ik} = \delta_{jk}$

Die $\{U_i\}$ sind alle lineare Summen von u.i.v. Normal ZV, und deshalb sind sie auch normalverteilt:

$$E[U_i] = 0$$

$$V[U_i] = 1$$

$$E[U_i U_j] = 0 \quad i \neq j$$

$$\begin{aligned}
\sum_{i=1}^{n} (X_i - \bar{X})^2 &= \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\
&= \sum_{i=1}^{n} U_i^2 - U_n^2 \\
&= \sum_{i=1}^{n-1} U_i^2
\end{aligned}$$

Da die $\{U_i\}$ u.i.v. $N(0,1)$ sind, folgt

$$(n-1)s^2 \sim \chi^2_{n-1}$$

16.5 Die Verteilung von s^2 für allgemeine $N(\mu, \sigma^2)$

Gegeben $\{X_i\}$ u.i.v. $N(\mu, \sigma^2)$

$$\Rightarrow \{Z_i\} \text{ u.i.v. } N(0,1) \text{ für } Z_i = \frac{X_i - \mu}{\sigma}$$

$$\bar{Z} \sim N(0, 1/n) \Rightarrow \bar{X} \sim N(0, \sigma^2/n)$$

$$\sum_{i=1}^{n} (Z_i - \bar{Z})^2 \sim \chi^2_{n-1}$$

$$\Rightarrow (n-1)s_Z^2 \sim \chi^2_{n-1}$$

$$\begin{aligned} s_X^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \\ &= \sigma^2 s_Z^2 \end{aligned}$$

$$\Rightarrow \frac{(n-1)s_X^2}{\sigma^2} \sim \chi^2_{n-1}$$

16.6 $\bar{X}$ und s^2

16.6.1 Die Unabhängigkeit von $\bar{X}$ und s^2 für $X_i \sim N(\mu, \sigma^2)$

Für $\{X_i\}$ u.i.v. N sind $\bar{X}$ und s^2 unabhängig. Student (W.S. Gosset) hat es erraten und Fisher hat es bewiesen. Die gemeinsame Dichte von den $\{U_i\}$ ZV, die durch die Helmert Transformation gewonnen worden sind, ist

$$f(u_1, \ldots, u_n) = ce^{-\frac{1}{2}\sum u_i^2}$$

da die $\{U_i\}$ u.i.v. $\sim N(0,1)$

$$= ce^{-\frac{1}{2}(\sum_{i=1}^{n-1} u_i^2 + u_n^2)}$$

$$= ce^{-\frac{1}{2}((n-1)s^2 + n\bar{x}_n^2)}$$

$$= g(s^2)h(\bar{x})$$

$\Rightarrow$ Unabhängigkeit

16.6.2 Die Verteilung von $\bar{X}$ und s^2 für $n = 2$

X_1, X_2 u.i.v. mit Dichte $g(x)$

$$f_{X_1 X_2}(x_1, x_2) = g(x_1)g(x_2)$$

$$\bar{X} = \frac{1}{2}(X_1 + X_2)$$

$$\begin{aligned} s^2 &= \frac{1}{2-1}\sum_{i=1}^{2}(X_i - \bar{X})^2 \\ &= \frac{1}{2}(X_1 - X_2)^2 \end{aligned}$$

Um die gemeinsame Verteilung von $\bar{X}$ und s^2 zu untersuchen, setzen wir

$$Y_1 = \frac{1}{2}(X_1 + X_2) = \bar{X}$$

$$Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2) = s$$

$$g_{\bar{X}s}(\bar{x}, s) = g(\bar{x} + \frac{s}{\sqrt{2}})g(\bar{x} - \frac{s}{\sqrt{2}})\sqrt{2}$$

weil $|J| = \sqrt{2}$
Offensichtlich sind $\bar{X}$ und s^2 im allgemeinen nicht unabhängig.

16.6.3 Beispiel für $\bar{X}$ und s^2

$$X \sim E(\lambda)$$

$$g(x) = \lambda^{-\lambda x}$$

$$g_{\bar{X}s}(\bar{x}, s) = \sqrt{2}\lambda^2 e^{-2\lambda\bar{x}}$$

$$0 \leq \bar{X} \quad \text{und} \quad |s| \leq \sqrt{2}\bar{X}$$

Die Marginaldichte von $\bar{X}$ ist

$$h(\bar{x}) = \int_{-\sqrt{2}\bar{x}}^{\sqrt{2}\bar{x}} g_{\bar{X}s}(\bar{x}, s)\mathrm{d}s$$

$$= \sqrt{2}\lambda^2 e^{-2\lambda\bar{x}} 2\sqrt{2}\bar{x}$$

$$= 4\lambda^2 \bar{x} e^{-2\lambda\bar{x}}$$

wie es natürlich sein sollte (da $\bar{X} = \frac{1}{2}(X_1 + X_2)$).

Obgleich $g_{\bar{X}s}(\bar{x}, s)$ s nicht enthält, sind $\bar{X}$ und s nicht unabhängig. (N.B. Wir haben hier mit $s = \pm\sqrt{s^2}$ gearbeitet, weder mit s^2, die Stichprobenvarianz, noch mit $|s|$, die Standardabweichung.)

16.7 Kleine Stichproben

Der ZGS ist für n groß gültig. Was kann man mit kleinen Stichproben machen?

Student zeigte, dass

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

eine t-Verteilung mit $(n-1)$ Freiheitsgraden hat, wenn

$$X_i \quad \text{u.i.v.} \quad \sim N(\mu, \sigma^2)$$

Die t-Verteilung hängt von $(n-1)$ ab. μ und σ spielen keine Rolle. Dadurch können wir Aussagen über T machen, sogar wenn beide Parameter der Verteilung, μ und σ, unbekannt sind.

16.8 Ableitung der t-Verteilung

Um die Dichte von

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

für $\{X_i\}$ u.i.v. $\sim N(\mu, \sigma^2)$ zu finden, benutzen wir die folgenden drei Resultate:

$$(1) \qquad \bar{X} \sim N(\mu, \sigma^2/n)$$

$$(2) \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

(3) $\bar{X}$ und s sind unabhängig.

Wir schreiben

$$T = \frac{\bar{Y}\sqrt{n}}{s_Y}$$

für

$$Y_i = \frac{X_i - \mu}{\sigma} \sim N(0, 1)$$

16.8.1 Transformationen für die t-Verteilung

$$R = \sqrt{n}\,\bar{Y}$$

$$U = (n-1)s_Y^2$$

$$V = \frac{U}{n-1} = s_Y^2$$

$$W = \sqrt{V} = s_Y$$

$$T = \frac{R}{W}$$

Für die letzte Transformation benutzen wir das Resultat:
X und Y unabhängig mit Dichten $f(x)$ und $g(y)$.
Für $Y > 0$ hat $T = \frac{X}{Y}$ die Dichte

$$h(t) = \int_0^\infty f(ty)g(y)y\mathrm{d}y$$

$$P(T \leq t) = \int\int_B f(x)g(y)\mathrm{d}x\mathrm{d}y$$

mit $B = \{(x,y) : x \leq yt\}$. Wir setzen $r = x/y$ und

$$P(T \leq t) = \int_{-\infty}^t \int_0^\infty f(ry)g(y)y\mathrm{d}y\mathrm{d}r = \int_{-\infty}^t h(r)\mathrm{d}r$$

16.8.2 Dichten für die t-Verteilung

$$f_R(r) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}r^2}$$

$$g_U(u) = \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}\, u^{\frac{n-1}{2}-1}\, e^{-\frac{u}{2}}$$

$$g_V(v) = \frac{n-1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}\, ((n-1)v)^{\frac{n-3}{2}}\, e^{-\frac{(n-1)v}{2}}$$

$$g_W(w) = \frac{(n-1)w}{2^{\frac{n-3}{2}}\Gamma(\frac{n-1}{2})}\, ((n-1)w^2)^{\frac{n-3}{2}}\, e^{-\frac{(n-1)w^2}{2}}$$

$$h_T(t) = \frac{1}{\sqrt{2\pi}}\frac{(n-1)^{\frac{n-1}{2}}}{2^{\frac{n-3}{2}}\Gamma(\frac{n-1}{2})}\int\limits_0^\infty e^{-\frac{1}{2}t^2y^2}y^{n-1}e^{-\frac{1}{2}(n-1)y^2}\mathrm{d}y$$

$$\Rightarrow h(t) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})}\,\frac{1}{\sqrt{n-1}}\,\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}$$

16.9 Verteilungen für Varianztests

Seien $V \sim \chi^2_m$ und $W \sim \chi^2_n$ unabhängig, dann kann man zeigen, dass

$$F = \frac{(V/m)}{(W/n)} \sim F_{m,n}$$

wo die Dichte einer $F_{m,n}$ Verteilung ist, für $x > 0$

$$h_{m.n}(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \, m^{\frac{m}{2}} \, n^{\frac{n}{2}} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}}$$

und $\quad h_{m.n}(x) = 0 \quad$ für $\quad x \leq 0$

Anwendungen

(1) Der Vergleich von zwei Varianzen

(2) Die Varianzanalyse