2. Repetition relevanter Teilbereiche der Statistik

(Maddala Kapitel 2)

Zufallsvariablen und Wahrscheinlichkeitsverteilungen

Zufallsvariable *X*
(stochastische Variable)

→ ***diskret:*** *X* kann nur eine endliche Anzahl von Werten annehmen

→ ***kontinuierlich:*** *X* kann (in einem bestimmten Intervall) jeden Wert annehmen

Realisation von *X*: *x*

Das Verhalten von Zufallsvariablen ist durch ihre ***Wahrscheinlichkeits-verteilung*** definiert. Diese wird für kontinuierliche Zufallsvariablen mit der sogenannten *Dichtefunktion* dargestellt (*probability density function, p.d.f.*).

Diskret (z.B. Münze): Prob(X = 0) = 0.5 Prob(X = 1) = 0.5

Kontinuierlich: *Dichtefunktion* *f(x)* mit den Eigenschaften

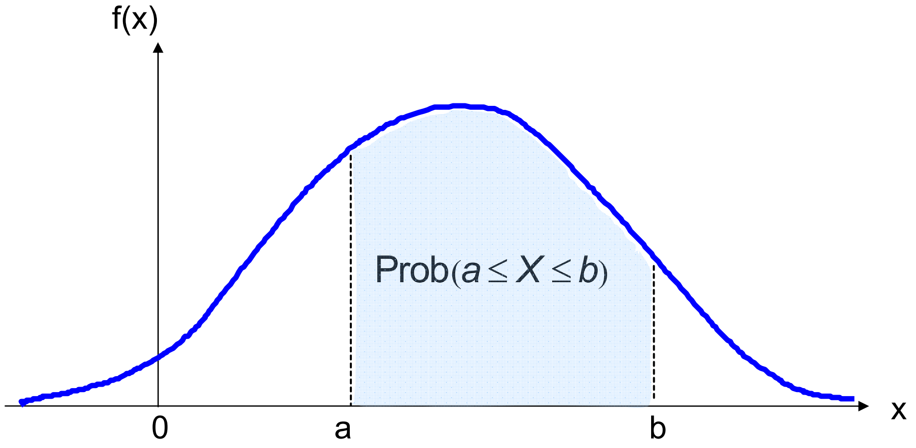
$$f(x) \geq 0 \text{ für alle } x \qquad \int_{-\infty}^{+\infty} f(x)\,dx = 1 \ .$$

Durch Integration lässt sich die Wahrscheinlichkeit bestimmen, dass X in ein bestimmtes Intervall [a,b] fällt:

$$\text{Prob}(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Kumulative Verteilungsfunktion: $F(c) = \text{Prob}(X \leq c) = \int_{-\infty}^{c} f(x)\,dx$

somit gilt: $\text{Prob}(a \leq X \leq b) = F(b) - F(a)$



Gemeinsame Verteilung von zwei (oder mehr) Zufallsvariablen

Im Fall von zwei (oder mehr) Zufallsvariablen ist zu unterscheiden zwischen der gemeinsamen Verteilung, der marginalen Verteilung (auch Randverteilung genannt) und der bedingten Verteilung:

- Die ***gemeinsame*** Verteilung von X und Y ist durch die Wahrscheinlichkeiten für alle möglichen Realisationen $X = x$ ***und*** $Y = y$ definiert.
- Die ***marginale*** Verteilung von X zeigt die Wahrscheinlichkeiten für die Realisationen $X = x$, egal welchen Wert Y annimmt (analog für Y).
- Die ***bedingte*** Verteilung von X zeigt die Wahrscheinlichkeiten für die Realisationen $X = x$, gegeben eine ***bestimmte*** Realisation $Y = y$ (analog für Y).

In Maddala, Seiten 18/19, findet sich ein Zahlenbeispiel für die gemeinsame Verteilung von zwei *diskreten* Zufallsvariablen, die voneinander unabhängig bzw. abhängig sind.

Im Fall von zwei *kontinuierlichen* Zufallsvariablen X und Y ist die Verteilung durch die ***gemeinsame*** Dichtefunktion $f(x,y)$ gegeben.

Die ***marginalen*** Dichtefunktionen von X bzw. Y erhält man durch Integration von $f(x,y)$ über y bzw. x:

$$f(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy \qquad\qquad f(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$$

(Intuitiv bedeutet z.B. Integration über y: Man lässt alle Werte von y zu.)

Die ***bedingten*** Verteilungen sind durch die Dichtefunktionen

$f(x|y)$ (Verteilung von X gegeben eine bestimmte Realisation $Y = y$) bzw.

$f(y|x)$ (Verteilung von Y gegeben eine bestimmte Realisation $X = x$)

definiert.

Die gemeinsame Dichtefunktion kann stets als Produkt von marginaler und bedingter Dichtefunktion geschrieben werden:

$$f(x,y) = f(x|y)\,f(y) = f(y|x)\,f(x)$$

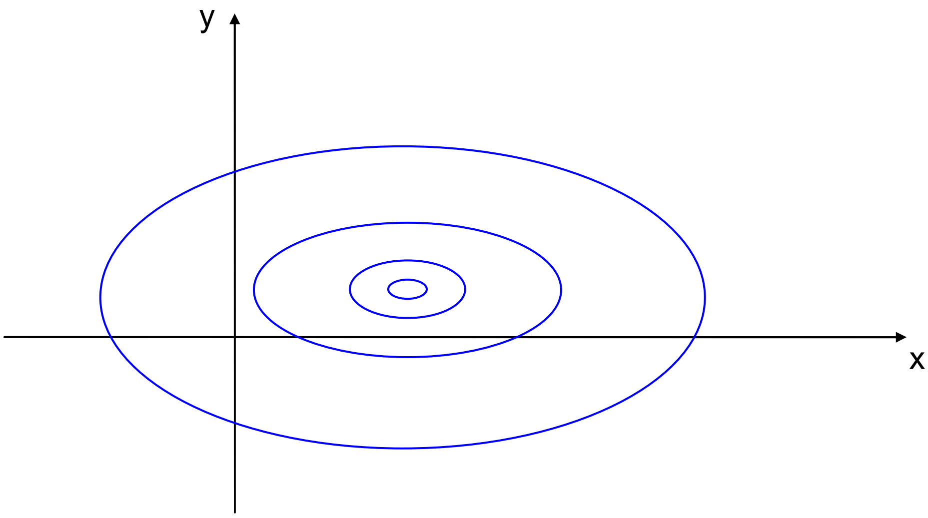
Falls die beiden Variablen voneinander ***unabhängig*** sind, gilt:

$f(x,y) = f(x)\, f(y)$ für alle x und y .

In diesem Fall sind die bedingten Dichtefunktionen gleich den marginalen Dichtefunktionen:

$f(x|y) = f(x)$ $\qquad$ $f(y|x) = f(y)$

Grafische Darstellung einer gemeinsamen Dichtefunktion $f(x,y)$ in Form von "Höhenkurven" eines "Dichtebergs", dessen Volumen über der (x,y)-Ebene gleich 1 ist:



Die Normalverteilung und verwandte Verteilungen

Normalverteilung

Dichtefunktion:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Erwartungswert: $E(x) = \mu$

Varianz: $E(x-\mu)^2 = \sigma^2$ $\qquad$ Standardabweichung: σ

Kurzschreibweise: $x \sim N(\mu, \sigma^2)$

Variablentransformation: $z = (x-\mu)/\sigma$

Die Variable z ist ***standard-normalverteilt***, d.h. $z \sim N(0,1)$.

Chi-Quadrat-Verteilung

Gegeben seien n unabhängig standard-normalverteilte Variablen:

$x_i \sim IN(0,1)$, $i = 1, ..., n$ (IN steht für **I**ndependent **N**ormal)

Eine daraus abgeleitete Variable

$$Z = \sum_i x_i^2$$

ist Chi-Quadrat-verteilt mit n Freiheitsgraden:

$$Z \sim \chi_n^2$$

Student's t-Verteilung

Gegeben sei: $x \sim IN(0,1)$ und $y \sim \chi_n^2$ (x und y unabhängig).

Eine daraus abgeleitete Variable

$$Z = x \big/ \sqrt{y/n}$$

ist t-verteilt mit n Freiheitsgraden:

$$Z \sim t_n$$

Für grosses n konvergiert t_n gegen eine Standard-Normalverteilung.
Für kleines n ist t_n flacher als die Standard-Normalverteilung.

F-Verteilung

Gegeben seien $y_1 \sim \chi_{n_1}^2$ und $y_2 \sim \chi_{n_2}^2$ (y_1 und y_2 unabhängig).

Eine daraus abgeleitete Variable

$$Z = \frac{y_1/n_1}{y_2/n_2}$$

ist F-verteilt mit n_1 und n_2 Freiheitsgraden:

$$Z \sim F_{n_1,n_2}$$

Die F-Verteilung ist somit die Verteilung des Verhältnisses von zwei voneinander unabhängig Chi-Quadrat-verteilten und durch die jeweilige Anzahl Freiheitsgrade geteilten Variablen.

Klassische deduktive Statistik

Datengenerierender Prozess (wahres Modell), charakterisiert durch *Wahrscheinlichkeits-Dichtefunktion* mit *unbekannten Parametern*:

z.B. $y_i \sim N(\mu,\sigma^2)$	z.B. $y_t = \alpha + \beta x_t + u_t$ mit $u_t \sim IN(0,\sigma^2)$
Parameter: μ, σ	Parameter: α, β, σ
Stichprobe:	
$y_1, y_2, \ldots, y_n$	$x_1, x_2, \ldots, x_T$ (exogen vorgegeben) $y_1, y_2, \ldots, y_T$
Schätzung:	
$\hat{\mu}, \hat{\sigma}$	$\hat{\alpha}, \hat{\beta}, \hat{\sigma}$

Unter einem sog. *Schätzer* versteht man eine Methode (eine Schätzformel), mit der man aus einer Stichprobe Schätzwerte für die Parameter eines datengenerierenden Prozesses ableitet. Für den ***Mittelwert*** *μ* einer normalverteilten Zufallsvariablen *y* lautet die Schätzformel:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Analog zu diesem Beispiel ist die ***Punktschätzung*** $\hat{\theta}$ eines unbekannten Parameters θ stets eine *Funktion der Stichprobenwerte* der Zufallsvariablen *y*:

$$\hat{\theta} = g(y_1, y_2, \ldots, y_n)$$

Dies gilt auch für die Schätzung eines ***Vertrauensintervalls:***

$$g_1(y_1, y_2, \ldots, y_n) \qquad g_2(y_1, y_2, \ldots, y_n) \qquad g_1 \le \theta \le g_2$$

Stichprobenverteilung

Die y_i sind *Zufallsvariablen* => Die geschätzten Parameter (z.B. $\hat{\mu}$ und $\hat{\theta}$) sind ebenfalls Zufallsvariablen, denn sie sind Funktionen von y_i. Ihre Wahrscheinlichkeitsverteilung wird ***Stichprobenverteilung*** genannt.

Die Eigenschaften von Schätzern leiten sich aus der Stichprobenverteilung der geschätzten Parameter ab. Wünschbare Eigenschaften sind:

- ***Unverzerrtheit***
- ***Effizienz***
- ***Konsistenz***

in ***kleinen*** Stichproben relevant

in ***grossen*** Stichproben relevant

Diese Eigenschaften charakterisieren die Beziehung zwischen den unbekannten Parametern des datengenerierenden Prozesses (θ) und den geschätzten Parametern ($\hat{\theta}$).

Unverzerrtheit (Erwartungstreue): $E(\hat{\theta}) = \theta$

Effizienz: Unter den unverzerrten Schätzern wird derjenige mit der kleinsten Varianz als effizient bezeichnet.

Konsistenz: Es sei $\hat{\theta}_n$ der Schätzer für θ aus einer Stichprobe vom Umfang *n*. Dieser Schätzer ist konsistent, falls gilt:

$$\lim_{n\to\infty}\left(\text{Prob}\left(|\,\hat{\theta}_n - \theta\,| < \varepsilon\right) = 1\right. \quad \text{für beliebig kleines } \varepsilon$$

In Worten locker formuliert: Ein Schätzer ist dann konsistent, wenn der Schätzwert mit wachsendem Stichprobenumfang immer näher zum unbekannten wahren Parameterwert rückt.

Stichprobenverteilung von Parameterschätzungen im Falle der Normalverteilung

Datengenerierender Prozess: $y \sim N(\mu, \sigma^2)$ $E(y) = \mu$
(bzw. Grundgesamtheit) $E(y-\mu)^2 = \sigma^2$

Stichprobe mit n Beobachtungswerten:

Schätzung μ: $\bar{y} = \frac{1}{n}\sum y_i$

Schätzung σ^2: $S^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$

Stichprobenverteilung von $\bar{y}$ und S^2:

$$\left.\begin{array}{l} \bar{y} \sim N(\mu, \sigma^2_{\bar{y}}) \\ \sigma^2_{\bar{y}} = \sigma^2 / n \\ \sigma_{\bar{y}} = \sigma / \sqrt{n} \end{array}\right]$$

$\bar{y}$ ist **normalverteilt** und ein unverzerrter und konsistenter Schätzer von μ.

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum (y_i - \bar{y})^2}{\sigma^2} = \sum \left(\frac{y_i - \bar{y}}{\sigma} \right)^2 \sim \chi^2_{n-1}$$

(**Chi-Quadrat-verteilt**)

$$\left[\frac{y_i - \mu}{\sigma} \sim N(0, 1) \implies \sum \left(\frac{y_i - \mu}{\sigma} \right)^2 \sim \chi^2_n \right]$$

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \implies \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}} = \frac{\bar{y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

(**t-verteilt**)

Hypothesentest

1. Eine statistische Hypothese ist eine ***Aussage über Parameterwerte eines datengenerierenden Prozesses*** (bzw. einer Grundgesamtheit). Sie wird anhand einer ***Stichprobe*** überprüft.
2. Wir unterscheiden zwischen ***Punkthypothesen*** (z.B. $\mu = 4$) und ***Intervall-hypothesen*** (z.B. $\mu > 0$).
3. Ein ***Hypothesentest*** ist ein Verfahren, das angibt, ob die Abweichung eines geschätzten Parameters von dem in der Hypothese postulierten Wert ***zufällig*** oder ***echt*** ist (z.B. $\bar{y} = 5$ vs. $\mu = 4$).
4. Die getestete Hypothese wird ***Null-Hypothese*** genannt und mit H_0 bezeichnet. Die ***alternative Hypothese*** wird mit H_1 bezeichnet.
5. Die Masszahl, welche die Abweichung der Parameterschätzung von dem in H_0 postulierten Wert angibt, wird ***Teststatistik*** genannt. Bei jedem Hypothesetest muss man für die Teststatistik einen ***kritischen Wert*** und ein entsprechendes ***Signifikanzniveau*** festlegen (häufig wird 5% gewählt). Dabei handelt es sich um die Wahrscheinlichkeit dafür, dass die berechnete Teststatistik bei Gültigkeit von H_0 den kritischen Wert über-schreitet und H_0 somit fälschlicherweise abgelehnt wird (was man auch als “Fehler erster Art“ bezeichnet).
6. Das ***beobachtete Signifikanzniveau*** (auch P-Wert genannt) ist die Wahrscheinlichkeit dafür, dass die Teststatistik bei Gültigkeit von H_0 den berechneten oder einen extremeren Wert annimmt. Gebräuliche Praxis:
   - P-Wert kleiner als 5%: Die Abweichung zwischen der Parameterschätzung und dem in H_0 postulierte Wert wird als statistisch ***signifikant*** bezeichnet, weil es unwahrscheinlich erscheint, dass eine Abweichung dieses Ausmasses durch Zufälligkeiten in der Stichprobe verursacht wurde. ***H_0 wird verworfen***.
   - P-Wert grösser als 5%: Die Abweichung zwischen der Parameterschätzung und dem in H_0 postulierte Wert wird als statistisch ***insignifikant*** bezeichnet, weil es nicht allzu unwahr-scheinlich ist, dass eine Abweichung dieses Ausmasses durch Zufälligkeiten in der Stichprobe verursacht wurde. ***H_0 wird nicht verworfen.***

Einfache Beispiele für Hypothesentests

1. Mittelwert einer normalverteilten Variablen

Datengenerierender Prozess: $y \sim N(\mu, \sigma^2)$

Nullhypothese H_0: $\mu = 0$ vs. Alternativhypothese H_1: $\mu \neq 0$

Stichprobe $n = 4$, $y_i = 1, -1, 3, 5$

geschätzter Mittelwert: $\bar{y} = \frac{1}{n}\sum y_i$ = (1-1+3+5)/4 = 2

geschätzte Varianz: $S^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$ = (1+9+1+9)/3 = 20/3 = 6.67

geschätzte Standardabweichung: $S = 2.58$

Standardabweichung der Mittelwertschätzung: $\hat{\sigma}_{\bar{y}} = S/\sqrt{n}$ = 1.29

berechneter t-Wert: $t_{ber} = \frac{\bar{y} - \mu}{S/\sqrt{n}}$ = 2/1.29 = 1.55

kritischer t-Wert (5%-Niveau, 2-seitiger Test): $t^* = 3.182$

$|t_{ber}| < t^*$ ⇨ H_0 nicht verwerfen !

P-Wert: $Prob(|t| \geq 1.55) = 0.22$ > 0.05 ⇨ H_0 nicht verwerfen !

2. Münzwurf

H_0: Münze ist regulär, dh. $P_{KOPF} = P_{ZAHL} = 0.5$

H_1: Münze ist nicht regulär

Testverfahren:

Die Münze wird 4 Mal geworfen. *Verwerfungskriterium:* Falls die Münze 4 mal auf Kopf *oder* 4 mal auf Zahl fällt, soll die Nullhypothese verworfen werden.

Unter der Nullhypothese gilt:

Prob(K,K,K,K) = $0.5^4 = 0.0625$

Prob(Z,Z,Z,Z) = $0.5^4 = 0.0625$

Mit diesem Verwerfungskriterium führt man folglich einen Test mit einem Signifikanzniveau von 12.5% durch, d.h. es besteht eine Wahrscheinlichkeit von 12.5%, dass eine richtige Nullhypothese fälschlicherweise verworfen wird (*Fehler erster Art*).

Reduziert man das Signifikanzniveau z.B. auf 5%, so wäre auch das Ereignis (K,K,K,K,K) oder (Z,Z,Z,Z,Z) mit einem beobachteten P-Wert von 6.25% nicht genügend unwahrscheinlich, um die Nullhypothese zu verwerfen. Bei diesem Vorgehen ergibt sich eine kleinere Wahrscheinlichkeit dafür, dass eine falsche Nullhypothese als falsch erkannt und verworfen wird (geringere *Macht* des Tests).