KUMULATIVE VERTEILUNGSFUNKTIONEN

von Thomas Shilgalis, Illinois State University, Normal,
Illinois 61761, USA
Originaltitel in "Teaching Statistics" Vol.10 (1988) Nr.3:
Cumulative Distribution Functions
Übertragung: Bernd Wollring, Universität Münster

Ist für eine Zufallsgröße X die Verteilungsdichte gegeben, so ist die kumulative Verteilungsfunktion durch die Gleichung $F(x) = P(X \leq x)$ definiert.

Ist X eine diskrete Zufallsgröße, so gilt:

$$F(x) = \sum_{t \leq x} f(t)$$

F ist eine Treppenfunktion, deren Sprünge an den Stellen t liegen, an denen $f(t) > 0$ gilt.

Ist X eine stetige Zufallsgröße, so gilt:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

und F ist stetig. Ferner gilt mit Ausnahme höchstens endlich vieler Punkte $F'(x) = f(x)$.

In jedem Fall ist F eine monoton wachsende Funktion, die rechtsseitig stetig ist.

Für diskrete Zufallsgrößen kann man f aufgrund der Gleichung $f(x) = P(X=x)$ leicht verstehen. Dagegen gilt für stetige Zufallsgrößen $P(X=x) = 0$ für jedes spezielle x, daher ist die Bedeutung von f schwerer zu erfassen. Im allgemeinen koppeln wir f und die Wahrscheinlichkeit eines Ereignisses A mit Hilfe von:

$$P(X \in A) = \int_A f(t)\,\mathrm{d}t$$

Ist A ein Intervall (a;b) , so haben wir die leicht verständliche Beziehung:

$$P(a \le X \le b) = \int_a^b f(t)\,dt = F(b) - F(a)$$

Die Bedeutung kumulativer Verteilungsfunktionen steht für Lehrende der Wahrscheinlichkeitsrechnung und Statistik außer Frage. Es ist jedoch ein Gegenstand, der Schülern und Studenten nicht einfach nahezubringen ist. Ich finde den im Folgenden skizzierten Weg hilfreich. Im verbleibenden Teil des Textes soll dargestellt werden, wie man mit Hilfe computererzeugter Werte einer Zufallsgröße X das Verstehen kumulativer Verteilungsfunktionen verbessern kann, indem man die Graphen der empirisch gefundenen und der theoretischen kumulativen Verteilungsfunktion miteinander vergleicht. Es werden drei Beispiele angegeben, ein diskretes und zwei mit stetiger Verteilungsdichte. (Eine Kopie oder Auflistung der entsprechenden Programme können beim Autor angefordert werden.)

Beispiel 1: Die Zufallsgröße X sei binomial verteilt mit Einzelchance p = 1/3 . Wir wählen n = 6 als Anzahl der Einzelversuche. Die Verteilungsdichte ist:

$$f(x) = \binom{6}{x}\left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x} \qquad x = 0,1,2,3,4,5,6$$

Die kumulative Verteilungsfunktion dazu ist:

$$F(x) = \sum_{t \le x} \binom{6}{t}\left(\frac{1}{3}\right)^t \left(\frac{2}{3}\right)^{6-t}$$

Für $x < 0$ ist $F(x) = 0$. Ist $x \geqslant 6$, so ist $F(x) = 1$.

Mit einem Computer erzeugen wir zwanzig Serien mit je sechs Versuchen. Die empirische Wahrscheinlichkeit F (2) zum Beispiel ist dann:

$$\frac{\text{Anzahl der Serien zu 6 Versuchen mit höchstens 2 Treffern}}{20}$$

Dieser Wert ist natürlich derselbe wie F (2.3) oder F (2.99) , die empirische kumulative Verteilungsfunktion ist daher genau wie die theoretische eine Treppenfunktion. Tabelle 1 zeigt die Ergebnisse eines Durchlaufs auf dem Computer.

x	*Number of sets of 6 trials with x or fewer successes*	*F*(x)*	*F(x)*
0	2	0·10	0·0878
1	8	0·40	0·3512
2	14	0·70	0·6804
3	19	0·95	0·8999
4	20	1·00	0·9821
5	20	1·00	0·9986
6	20	1·00	1·0000

Tabelle 1

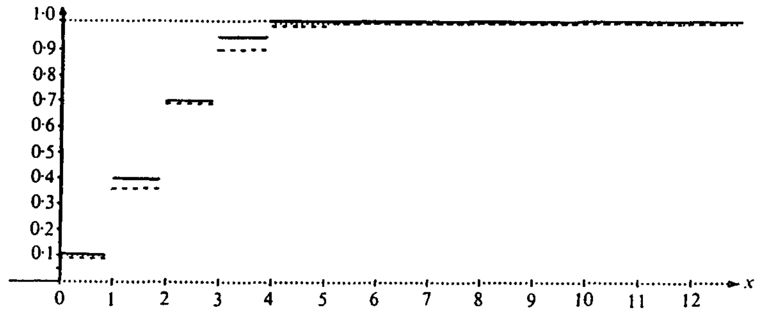

Bild 1 : F(x) für die Binominalverteilung mit p = 1/3 , n = 6 Einzelversuchen in 20 Versuchsserien

Bild 1 zeigt die Graphen von F* und F, wobei durchgezogene Linien F* und gepunktete F bezeichnen.

Beispiel 2: X besitze eine Gleichverteilung im Intervall (2;8). Die Wahrscheinlichkeitsdichte ist:

$$f(x) = 1/6 \qquad 2 \leq x \leq 8$$

Die kumulative Verteilungsfunktion ist:

$$F(x) = \begin{cases} 0 & x < 2 \\ (x-2)/6 & 2 \leq x \leq 8 \\ 1 & x > 8 \end{cases}$$

Zur Erzeugung beobachteter Werte von X benötigen wir eine Transformation, da der Zufallszahlen-Generator in unserem Computer (APPLE) Werte erzeugt, die im Intervall (0;1) gleich verteilt sind. Bezeichnet Y die entsprechende Zufallsgröße, so benutzen wir x = 6y+2 .

Die empirische kumulative Verteilungsfunktion F erhalten wir wie im folgenden Beispiel:

$$F(4.6) = \frac{\text{Zahl der Beobachtungen} \leqslant 4.6}{20}$$

Man beachte, daß jeder beobachtete Wert von X zwischen 2 und 8 auftreten kann, aber bei jedem Durchlauf des Computers höchstens 20 verschiedene Werte auftreten, so daß F^* wieder eine Treppenfunktion ist, obwohl X eine stetige Zufallsgröße ist. Dagegen ist F stetig. Die Ergebnisse eines Durchlaufs im Computer zeigt Tabelle 2. Bild 2 zeigt die Graphen von F und F^*, wobei zur Darstellung von F^* Streckenabschnitte und für F eine Kurve Anwendung finden.

Tabelle 2

x	*Number of values* $\leq x$	$F^*(x)$	$F(x)$
2·40	1	0·05	0·067
2·52	2	0·10	0·087
2·84	4	0·20	0·140
3·03	5	0·25	0·172
3·33	6	0·30	0·222
4·28	7	0·35	0·380
4·55	8	0·40	0·425
4·95	9	0·45	0·492
4·98	10	0·50	0·497
5·05	11	0·55	0·508
5·48	12	0·60	0·580
5·97	13	0·65	0·662
6·33	14	0·70	0·722
6·90	15	0·75	0·817
6·92	16	0·80	0·820
7·06	17	0·85	0·843
7·20	18	0·90	0·867
7·55	19	0·95	0·925
7·92	20	1·00	0·987

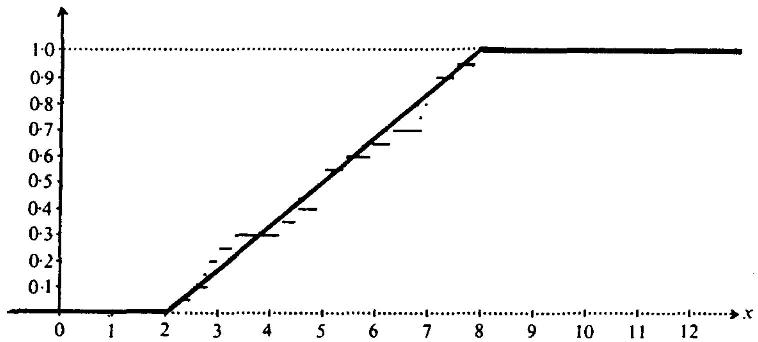

Bild 2 : $F(x) = (x-2)/6$; $2 \leqslant x \leqslant 8$; 20 Versuche

Beispiel 3 : X habe eine Exponentialverteilung mit Parameter 3. Die Verteilungsdichte ist:

$$f(x) = (\tfrac{1}{3})e^{-x/3} \qquad x > 0$$

Die kumulative Verteilungsfunktion ist:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x/3} & x \geq 0 \end{cases}$$

Auch hier benötigen wir eine Transformation zur Darstellung beobachteter Werte der Zufallsgröße X. Wir erzeugen mit dem APPLE-Computer wieder Werte der Zufallsgröße Y , die über (0;1) gleichverteilt ist. Die passende Transformation ist x = - 3 ln (y). Die Transformation erhalten wir aus:

$$F(x) = \mathrm{P}(X \leq x) = \mathrm{P}(-3 \ln Y \leq x) = \mathrm{P}(\ln Y \geq -x/3)$$
$$= \mathrm{P}(Y \geq e^{-x/3}) = 1 - \mathrm{P}(Y \leq e^{-x/3}) = 1 - e^{-x/3}$$

x	*No. of of values ≤ x*	*F*(x)*	*F(x)*	*x*	*No. of values ≤ x*	*F*(x)*	*F(x)*
0·25	1	0·05	0·080	3·22	11	0·55	0·658
0·26	2	0·10	0·083	3·70	12	0·60	0·709
0·42	3	0·15	0·131	3·78	13	0·65	0·716
1·40	4	0·20	0·373	4·01	14	0·70	0·737
1·80	5	0·25	0·451	5·31	15	0·75	0·830
1·87	6	0·30	0·464	5·37	16	0·80	0·833
2·51	7	0·35	0·567	5·39	17	0·85	0·834
2·54	8	0·40	0·571	5·42	18	0·90	0·836
2·61	9	0·45	0·581	6·19	19	0·95	0·872
2·78	10	0·50	0·604	7·60	20	1·00	0·920

Tabelle 3

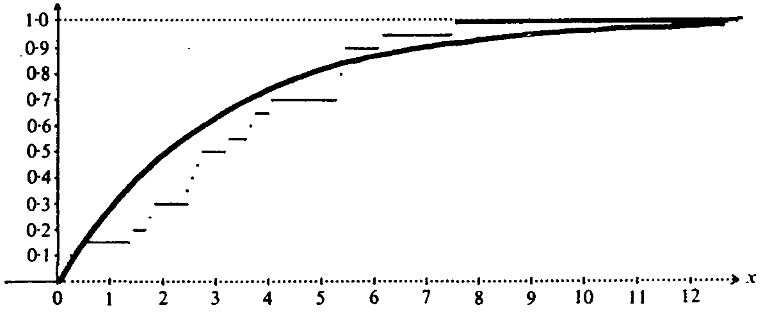

Bild 3 : F(x) = (1/3)exp(-x/3) ; 20 Versuche

Die letzte Gleichung folgt aus der Tatsache, daß F(y) = y gilt, wenn Y auf (0;1) gleichverteilt ist.

Tabelle 3 zeigt die Ergebnisse eines Durchlaufs auf dem Computer. Bild 3 zeigt die Graphen von F und F*, wobei F* mit Streckenabschnitten und F mit einer Kurve dargestellt ist.

Wenn man das Programm, das diese Serien erzeugt, in der Klasse einsetzt, so sollte man jeden Durchlauf stets zwei- oder dreimal wiederholen, wobei die Anzahl der Beobachtungen konstant bleibt. Dies zeigt, daß die empirische kumulative Verteilungsfunktion sich bei jedem Durchlauf ändert, während die theoretische dies nicht tut. Es ist ferner wichtig, langsam genug vorzugehen, so daß die Lernenden sehen, wie die empirische kumulative Verteilungsfunktion zustande kommt. Es ist zudem instruktiv, die Anzahl der Beobachtungen zu variieren. Erhöht man ihre Anzahl, so erhält man in der Regel eine bessere Übereinstimmung der theoretischen und der empirischen kumulativen Verteilungsfunktionen. Dies ist in Bild 4 dargestellt, wobei Beispiel 3 mit 40 Beobachtungen wiederholt wird und für F und F dieselben Darstellungen gewählt wurden.

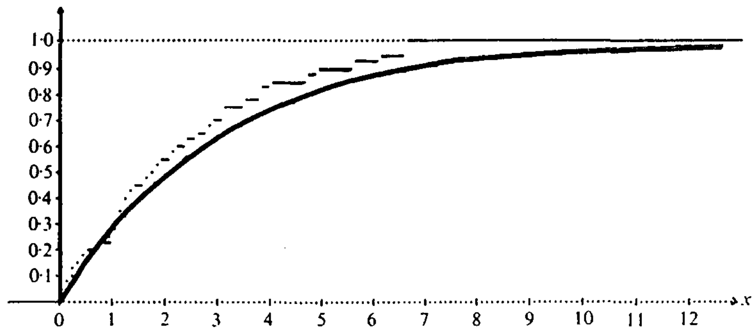

Bild 4 : F(x) = (1/3)exp(-x/3) ; 40 Versuche