

## 5 Die Poisson-Approximation

Im vierten Kapitel hatten wir mit der Normalverteilung die sicherlich wichtigste und meiststudierte Verteilung der W.-Theorie kennengelernt und gesehen, daß man diese als Limes einer geeignet skalierten Binomialverteilung erhalten kann. In diesem Kapitel werden wir eine weitere zentrale Verteilung kennenlernen, die sich ebenfalls als Limes einer (natürlich anders skalierten) Binomialverteilung schreiben läßt.

Wir wollen diese Verteilung an einem Beispiel kennenlernen.

### *Das Experiment von Rutherford und Geiger*

In einem bekannten Experiment beobachteten die Physiker Rutherford und Geiger den Zerfall einer radioaktiven Substanz. Genauer studierten sie die Emission von  $\alpha$ -Teilchen eines radioaktiven Präparates in  $n = 2608$  Zeitabschnitten von 7.5 Sekunden. Die folgende Tabelle gibt die Versuchsergebnisse wieder. Hierbei steht  $n_i$  für jedes natürliche  $i$  für die Anzahl der Zeitabschnitte, in denen genau  $i$   $\alpha$ -Teilchen emittiert wurden,  $r_i$  bezeichnet die relativen Häufigkeiten dieser Zeitabschnitte.

$i$	$n_i$	$r_i$
0	57	0.02186
1	203	0.0778
2	383	0.1469
3	525	0.2013
4	532	0.2040
5	408	0.1564
6	273	0.1047
7	139	0.0533
8	45	0.0173
9	27	0.0103
10	10	0.0038
11	4	0.0015
12	0	0
13	1	0.0004
14	1	0.0004

Offensichtlich sind diese Daten weit davon entfernt von einer Normalverteilung zu stammen. Wir benötigen vielmehr eine Verteilung, die die "Enden", d.h. die großen Zahlen mit einem sehr viel kleineren Gewicht versieht. Eine solche Verteilung ist die *Poisson-Verteilung*.

**(5.1) Definition.** Sei  $\lambda > 0$  eine reelle Zahl. Eine Zufallsgröße  $X$  mit  $X(\Omega) = \mathbb{N}_0$  und der Verteilung  $\pi_\lambda$  gegeben durch

$$\pi_\lambda(k) = \frac{e^{-\lambda}}{k!} \lambda^k, \quad k \in \mathbb{N}_0,$$

heißt *Poisson-verteilt mit Parameter  $\lambda > 0$* .

Zunächst bemerken wir, daß die Poisson-Verteilung auf den natürlichen Zahlen, incl. der Null  $\mathbb{N}_0$  konzentriert ist. Desweiteren überzeugt man sich rasch, daß

$$\sum_{k=0}^{\infty} \pi_{\lambda}(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

ist.  $\pi_{\lambda}$  ist also tatsächlich eine Wahrscheinlichkeit.

Der Erwartungswert dieser Verteilung ist leicht zu berechnen:

$$\sum_{k=0}^{\infty} k \pi_{\lambda}(k) = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

Eine Poisson-verteilte Zufallsgröße hat also Erwartungswert  $\lambda$ .

Als nächstes wollen wir die Varianz ausrechnen:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \pi_{\lambda}(k) = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} (k(k-1) + k) \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+2}}{k!} + \lambda = \lambda^2 + \lambda. \end{aligned}$$

Somit gilt

$$V(X) = E(X^2) - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Wir fassen diese beiden Feststellungen noch einmal in folgendem Lemma zusammen.

**(5.2) Lemma.** *Erwartungswert und Varianz einer Poisson-verteilten Zufallsgröße sind gleich dem Parameter  $\lambda$ .*

Wir wollen nun einmal die eingangs gezeigten Daten aus Rutherford's Experiment mit denen einer Poissonverteilung vergleichen. Dabei stellt sich die Frage, wie wir den Parameter  $\lambda$  am geschicktesten wählen. Vor dem Hintergrund des Gesetzes der großen Zahlen, nach dem man eine mittlere Zahl emittierter Teilchen erwarten kann, die nahe am Erwartungswert liegt und Lemma (5.2) ist eine gute Wahl die,  $\lambda$  als die durchschnittliche Anzahl der Emissionen zu wählen. Diese betrug im Experiment von Rutherford und Geiger

$$a = \frac{10097}{2608} \sim 3.87.$$

Die nächste Tabelle zeigt den Vergleich der relativen Häufigkeiten  $r_k$  aus dem Experiment von Rutherford und Geiger mit den Wahrscheinlichkeiten  $\pi_{\lambda}(k)$  einer Poissonverteilung zum Parameter  $\lambda = 3.87$ .

$k$	$r_k$	$\pi_\lambda(k)$
0	0.0219	0.0208
1	0.0778	0.0807
2	0.1469	0.1561
3	0.2013	0.2015
4	0.2040	0.1949
5	0.1564	0.1509
6	0.1047	0.0973
7	0.0533	0.0538
8	0.0173	0.0260
9	0.0103	0.0112
10	0.0038	0.0043
11	0.0015	0.0015
12	0	0.0005
13	0.0004	0.0002
14	0.0004	$4 \times 10^{-5}$

Die beobachteten relativen Häufigkeiten differieren also von den durch die entsprechende Poisson-Verteilung vorhergesagten Werten nur um wenige Tausendstel. Warum dies ein plausibles Ergebnis ist, soll am Ende dieses Kapitels in einem Satz geklärt werden, der zeigen wird, daß viele Prozesse, die einer Reihe von Anforderungen genügen, eine Poisson-Approximation erlauben. Grundlage dieses Satzes ist eine Feststellung darüber, wie genau sich die Binomialverteilung  $b(\cdot; n, p)$  für kleine Parameter  $p$  und große  $n$  durch die Poissonverteilung  $\pi_\lambda(k)$  approximieren läßt. Wieder bleibt das Problem,  $\lambda$  zu wählen. Wir lösen es so, daß wir  $\lambda$  so bestimmen, daß die Erwartungswerte der Binomialverteilung und der Poissonverteilung übereinstimmen, daß also  $\lambda = np$  ist. Wir wollen also zeigen:  $b(k; n, p)$  liegt nahe bei  $\pi_\lambda(k)$  für  $\lambda = np$ .

Um das zu präzisieren, benötigen wir ein Maß für den Abstand zweier Wahrscheinlichkeiten. Dies wird in unserem Fall gegeben sein durch

$$\Delta(n, p) := \sum_{k=0}^{\infty} |b(k; n, p) - \pi_{np}(k)|.$$

$\Delta(n, p)$  läßt sich ähnlich auch für den Abstand beliebiger anderer Wahrscheinlichkeiten definieren und heißt *Abstand der totalen Variation*.

Wir zeigen das folgende Resultat, das sogar noch wesentlich weitreichender ist als unser oben gestecktes Ziel:

**(5.3) Satz.** Es seien  $X_1, \dots, X_n$  unabhängige Zufallsvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum, mit  $P(X_i = 1) = p_i$  und  $P(X_i = 0) = 1 - p_i$  mit  $0 < p_i < 1$  für alle  $i = 1, \dots, n$ . Sei  $X = X_1 + \dots + X_n$  und  $\lambda = p_1 + \dots + p_n$ , dann gilt:

$$\sum_{k=0}^{\infty} |P(X = k) - \pi_\lambda(k)| \leq 2 \sum_{i=1}^n p_i^2.$$

Es folgt also im Fall  $p = p_1 = \dots = p_n$ :

**(5.4) Satz.** Für alle  $n \in \mathbb{N}$  und  $p \in (0, 1)$  gilt  $\Delta(n, p) \leq 2np^2$ .

Die Schranken in den Sätzen (5.3) und (5.4) sind natürlich nur interessant, falls  $\sum_{i=1}^n p_i^2$  klein wird bzw.  $p^2$  klein wird gegen  $n$ . Offenbar benötigt man in Satz (5.4) dazu mindestens  $p \ll \frac{1}{\sqrt{n}}$ , d.h. die Wahrscheinlichkeit eines Einzelerfolges wird klein mit  $n$ . Aus diesem Grund heißt die Poisson-Verteilung auch Verteilung seltener Ereignisse. Insbesondere folgt der sogenannte *Poissonsche Grenzwertsatz*, der von *Siméon Denis Poisson* (1781-1840) im Jahre 1832 entdeckt wurde:

**(5.5) Satz.** (*Grenzwertsatz von Poisson*) Ist  $\lambda > 0$  und gilt  $np_n \rightarrow \lambda > 0$  für  $n \rightarrow \infty$ , so gilt für jedes  $k \in \mathbb{N}_0$ :

$$\lim_{n \rightarrow \infty} b(k; n, p_n) = \pi_\lambda(k).$$

(5.5) folgt sofort aus (5.4): Aus  $np_n \rightarrow \lambda$  folgt  $p_n \rightarrow 0$  für  $n \rightarrow \infty$  und  $np_n^2 \rightarrow 0$ . Ferner ist  $|b(k; n, p) - \pi_{np}(k)| \leq \Delta(n, p)$  für jedes  $k \in \mathbb{N}_0$ . Demzufolge gilt

$$\lim_{n \rightarrow \infty} |b(k; n, p_n) - \pi_{np_n}(k)| = 0.$$

Wegen  $\pi_{np_n}(k) \rightarrow \pi_\lambda(k)$  folgt (5.5).

Offenbar unterscheidet sich (5.4) von (5.5) dadurch, daß die Aussage von (5.4) auch im Fall, wo  $np_n^2 \rightarrow 0$ ,  $np_n \rightarrow \infty$  gilt, von Interesse ist (z.B.  $p_n = 1/n^{2/3}$ ). Der wichtigste Vorzug von (5.3) und (5.4) im Vergleich zu (5.5) ist jedoch, daß eine ganz konkrete Approximationsschranke vorliegt. Dafür ist Satz (5.3) auch schwieriger zu beweisen als (5.5) (den wir hier allerdings nur als Korollar aus Satz (5.4) ableiten wollen).

Bevor wir den Beweis von Satz (5.3) geben, stellen wir einen wichtigen Aspekt der Poissonverteilung bereit:

**(5.6) Proposition.**  $X$  und  $Y$  seien unabhängig und Poisson-verteilt mit Parametern  $\lambda$  beziehungsweise  $\mu > 0$ . Dann ist  $X + Y$  Poisson-verteilt mit Parameter  $\lambda + \mu$ .

*Beweis.* Für  $n \in \mathbb{N}_0$  gilt:

$$\begin{aligned} P(X + Y = n) &= \sum_{k=0}^n P(X = k, Y = n - k) \\ &= \sum_{k=0}^n P(X = k)P(Y = n - k) \quad (\text{Unabhängigkeit}) \\ &= \sum_{k=0}^n \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} e^{-\lambda} e^{-\mu} = \frac{1}{n!} \left( \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \right) e^{-(\lambda+\mu)} \\ &= \frac{1}{n!} (\lambda + \mu)^n e^{-(\lambda+\mu)} = \pi_{\lambda+\mu}(n). \end{aligned}$$

□

**(5.7) Bemerkung.** Per Induktion folgt sofort, daß die Summe von endlich vielen unabhängigen Poisson-verteilten Zufallsgrößen wieder Poisson-verteilt ist, wobei der Parameter sich als Summe der Einzelparameter ergibt.

*Beweis von Satz 5.3.*

Der Beweis des Satzes (5.3) verwendet eine Technik, die man *Kopplung* (*coupling*) nennt.

Dabei verwenden wir wesentlich, daß bei der Berechnung des Abstands

$\sum_{k=0}^{\infty} |P(X = k) - \pi_{\lambda}(k)|$  die Größen  $P(X = k)$  bzw.  $\pi_{\lambda}(k)$  zwar die Verteilungen von Zufallsvariablen sind, daß aber in die Berechnung der zugrunde liegende W.-Raum nicht eingeht. Wir können also einen W.-Raum und Zufallsvariablen mit den gegebenen Verteilungen so wählen, daß sie für unsere Zwecke besonders geeignet sind und das bedeutet, daß sie sich bei gegebener Verteilung möglichst wenig unterscheiden. Konkret konstruieren wir:

Sei  $\Omega_i = \{-1, 0, 1, 2, \dots\}$ ,  $P_i(0) = 1 - p_i$  und  $P_i(k) = \frac{e^{-p_i}}{k!} p_i^k$  für  $k \geq 1$  sowie  $P_i(-1) = 1 - P_i(0) - \sum_{k \geq 1} P_i(k) = e^{-p_i} - (1 - p_i)$ . Nach Konstruktion sind somit  $(\Omega_i, P_i)$  W.-Räume. Betrachte dann den Produktraum  $(\Omega, P)$  der  $(\Omega_i, P_i)$  im Sinne der Definition (2.13). Wir setzen für  $\omega \in \Omega$

$$X_i(\omega) := \begin{cases} 0, & \text{falls } \omega_i = 0, \\ 1, & \text{sonst,} \end{cases}$$

und

$$Y_i(\omega) := \begin{cases} k, & \text{falls } \omega_i = k, k \geq 1, \\ 0, & \text{sonst.} \end{cases}$$

Dann haben nach Definition die Zufallsgrößen  $X_i$  die geforderte Verteilung:  $P(X_i = 1) = p_i$  und  $P(X_i = 0) = 1 - p_i$ . Sie sind weiter nach Definition des Produktraumes unabhängig. Die  $Y_i$  sind nach Definition Poisson-verteilt zum Parameter  $p_i$  und ebenfalls unabhängig. Also folgt mit Proposition (5.6), daß  $Y = Y_1 + \dots + Y_n$  Poisson-verteilt ist zum Parameter  $\lambda$ . Nun stimmen die Zufallsgrößen in den Werten 0 und 1 überein, und es ist  $P(X_i = Y_i) = P_i(0) + P_i(1) = (1 - p_i) + e^{-p_i} p_i$ , und somit

$$P(X_i \neq Y_i) = p_i(1 - e^{-p_i}) \leq p_i^2,$$

denn für  $x > 0$  gilt  $1 - e^{-x} \leq x$ . Damit folgt

$$\begin{aligned} & \sum_{k=0}^{\infty} |P(X = k) - \pi_{\lambda}(k)| = \sum_{k=0}^{\infty} |P(X = k) - P(Y = k)| \\ &= \sum_{k=0}^{\infty} |P(X = k = Y) + P(X = k \neq Y) - (P(X = k = Y) + P(X \neq k = Y))| \\ &\leq \sum_{k=0}^{\infty} P(X = k \neq Y) + P(X \neq k = Y) \\ &= 2P(X \neq Y) \leq 2 \sum_{i=1}^n P(X_i \neq Y_i) \leq 2 \sum_{i=1}^n p_i^2. \end{aligned}$$

Das beweist Satz (5.3). □

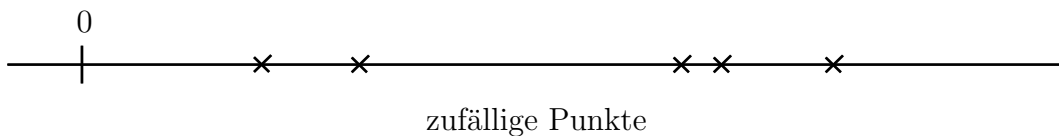
Nun können wir auch klären, warum die Ergebnisse im Experiment von Rutherford und Geiger so erstaunlich nahe an den Vorhersagen einer Poisson-Verteilung lagen. Dies geschieht im Rahmen des sogenannten Poissonschen Punktprozesses.

*Der Poissonsche Punktprozeß (Poisson point process)*

Wir konstruieren ein mathematisches Modell für auf einer Zeitachse zufällig eintretende Vorkommnisse. Beispiele sind etwa: Ankommende Anrufe in einer Telefonzentrale, Registrierung radioaktiver Teilchen in einem Geigerzähler, Impulse in einer Nervenfasern etc.

Die Zeitachse sei  $(0, \infty)$ , und die „Vorkommnisse“ seien einfach zufällige Punkte auf dieser Achse. Die Konstruktion eines unterliegenden Wahrscheinlichkeitsraumes ist leider etwas aufwendig und soll hier einfach weggelassen werden (wir glauben hier einfach mal, daß man das kann).

Ist  $I = (t, t + s]$  ein halboffenes Intervall, so bezeichnen wir mit  $N_I$  die zufällige Anzahl der Punkte in  $I$ .  $N_I$  ist also eine Zufallsgröße mit Werten in  $\mathbb{N}_0$ . Statt  $N_{(0,t]}$  schreiben wir auch einfach  $N_t$ .



An unser Modell stellen wir eine Anzahl von Bedingungen (P1) bis (P5), die für Anwendungen oft nur teilweise realistisch sind.

- (P1) Die Verteilung von  $N_I$  hängt nur von der Länge des Intervalls  $I$  ab. Anders ausgedrückt: Haben die beiden Intervalle  $I, I'$  dieselbe Länge, so haben die Zufallsgrößen  $N_I$  und  $N_{I'}$  dieselbe Verteilung. Man bezeichnet das auch als (zeitliche) Homogenität des Punktprozesses.
- (P2) Sind  $I_1, I_2, \dots, I_k$  paarweise disjunkte Intervalle, so sind  $N_{I_1}, N_{I_2}, \dots, N_{I_k}$  unabhängige Zufallsgrößen.
- (P3) Für alle  $I$  (stets mit endlicher Länge) existiert  $EN_I$ . Um Trivialitäten zu vermeiden, fordern wir:
- (P4) Es existiert ein Intervall  $I$  mit  $P(N_I > 0) > 0$ .

Aus (P1), (P3), (P4) lassen sich schon einige Schlüsse ziehen: Sei

$$\lambda(t) = EN_t \geq 0.$$

Offensichtlich gilt  $\lambda(0) = 0$ , denn  $N_0$  setzen wir natürlich 0. Die Anzahl der Punkte in einer Vereinigung disjunkter Intervalle ist natürlich die Summe für die Einzelintervalle. Insbesondere gilt:

$$N_{t+s} = N_t + N_{(t,t+s]}.$$

Demzufolge:

$$\lambda(t + s) = \lambda(t) + EN_{(t,t+s]},$$

was wegen (P1)

$$= \lambda(t) + \lambda(s)$$

ist.

Nach einem Satz aus der Analysis, der hier nicht bewiesen werden soll, muß eine derartige Funktion linear sein, das heißt, es existiert  $\lambda \geq 0$  mit  $\lambda(s) = \lambda s$ .  $\lambda = 0$  können wir wegen (P4) sofort ausschließen. In diesem Fall müßte nach (P1)  $EN_I = 0$  für jedes Intervall gelten. Dies widerspricht offensichtlich (P4).

Für kleine Intervalle ist die Wahrscheinlichkeit dafür, daß überhaupt ein Punkt in diesem Intervall liegt, klein. Es gilt nämlich:

$$P(N_I \geq 1) = \sum_{k=1}^{\infty} P(N_I = k) \leq \sum_{k=1}^{\infty} kP(N_I = k) = EN_I$$

und demzufolge

$$P(N_{(t,t+\varepsilon]} \geq 1) \leq \lambda\varepsilon \quad \text{für alle } t, \varepsilon \geq 0.$$

Unsere letzte Forderung besagt im wesentlichen, daß sich je zwei Punkte separieren lassen, es also keine Mehrfachpunkte gibt. Dazu sei für  $T > 0$

$$D_T(\omega) := \inf_{t,s \leq T} \{|t-s| : |N_t - N_s| \geq 1\}$$

dann besagt unsere Forderung (P5):

$$(P5) \quad P(D_T \leq \alpha_n) \xrightarrow{n \rightarrow \infty} 0$$

für jede Nullfolge  $\alpha_n$  und jedes endliche  $T$ .

Natürlich haben wir in keiner Weise belegt, daß eine Familie von Zufallsgrößen  $N_I$  mit den Eigenschaften (P1)–(P5) als mathematisches Objekt existiert. Wir können dies im Rahmen dieser Vorlesung nicht tun. Wir können jedoch nachweisen, daß für einen Punktprozeß, der (P1) bis (P5) erfüllt, die  $N_I$  alle Poisson-verteilt sein müssen:

**(5.8) Satz.** Sind (P1) bis (P5) erfüllt, so sind für alle  $t, s \geq 0$  die Zufallsgrößen  $N_{(t,t+s]}$  Poisson-verteilt mit Parameter  $\lambda s$ .

*Beweis.* Wegen (P1) genügt es,  $N_s = N_{(0,s]}$  zu betrachten. Wir halten  $s > 0$  fest. Für  $k \in \mathbb{N}$ ,  $1 \leq j \leq k$ , definieren wir

$$X_j^{(k)} := N_{(s(j-1)/k, sj/k]}$$

$$\bar{X}_j^{(k)} := \begin{cases} 1, & \text{falls } X_j^{(k)} \geq 1, \\ 0, & \text{falls } X_j^{(k)} = 0. \end{cases}$$

Für jedes feste  $k$  sind die  $X_j^{(k)}$  nach (P2) unabhängig und die  $\bar{X}_j^{(k)}$  damit ebenfalls.

Wir stellen einige einfach zu verifizierende Eigenschaften dieser Zufallsgrößen zusammen:

$$N_s = \sum_{j=1}^k X_j^{(k)}.$$

Sei  $\bar{N}_s^{(k)} := \sum_{j=1}^k \bar{X}_j^{(k)}$ . Dann gilt für jede mögliche Konfiguration der Punkte:

$$\bar{N}_s^{(k)} \leq N_s.$$

Demzufolge gilt für jedes  $m \in \mathbb{N}$ :

$$P(\bar{N}_s^{(k)} \geq m) \leq P(N_s \geq m). \quad (5.1)$$

Sei  $p_k = P(\bar{X}_i^{(k)} = 1) = P(X_i^{(k)} \geq 1) = P(N_{s/k} \geq 1)$ .

$$\bar{N}_s^{(k)} \text{ ist binomialverteilt mit Parameter } k, p_k. \quad (5.2)$$

Wir verwenden nun (P5), um nachzuweisen, daß sich für große  $k$   $\bar{N}_s^{(k)}$  nur wenig von  $N_s$  unterscheidet. In der Tat bedeutet ja  $\bar{N}_s^{(k)} \neq N_s$ , daß es mindestens ein Intervall der Länge  $1/k$  gibt, in dem 2 Punkte liegen, also

$$\{\bar{N}_s^{(k)} \neq N_s\} \subseteq \{D_s \leq 1/k\}.$$

Wegen (P5) folgt

$$P(\bar{N}_s^{(k)} \neq N_s) \leq P(D_s \leq 1/k) \rightarrow 0 \quad (5.3)$$

für  $k \rightarrow \infty$ . Für  $m \in \mathbb{N}$  und  $k \in \mathbb{N}$  gilt:

$$\begin{aligned} P(N_s = m) &\geq P(\bar{N}_s^{(k)} = m, \bar{N}_s^{(k)} = N_s) \\ &\geq P(\bar{N}_s^{(k)} = m) - P(\bar{N}_s^{(k)} \neq N_s) \\ P(N_s = m) &\leq P(\bar{N}_s^{(k)} = m, \bar{N}_s^{(k)} = N_s) + P(\bar{N}_s^{(k)} \neq N_s) \\ &\leq P(\bar{N}_s^{(k)} = m) + P(\bar{N}_s^{(k)} \neq N_s). \end{aligned}$$

Unter Benutzung von (5.2) und (5.3) folgt:

$$P(N_s = m) = \lim_{k \rightarrow \infty} P(\bar{N}_s^{(k)} = m) = \lim_{k \rightarrow \infty} b(m; k, p_k) \quad (5.4)$$

und analog

$$P(N_s \geq m) = \lim_{k \rightarrow \infty} P(\bar{N}_s^{(k)} \geq m). \quad (5.5)$$

Wir zeigen nun:

$$\lim_{k \rightarrow \infty} kp_k = \lambda s. \quad (5.6)$$

$$kp_k = E\bar{N}_s^{(k)} = \sum_{j=1}^{\infty} jP(\bar{N}_s^{(k)} = j) = \sum_{l=1}^{\infty} P(\bar{N}_s^{(k)} \geq l).$$

$P(\bar{N}_s^{(k)} \geq l)$  ist nach (5.1) nicht größer als  $P(N_s \geq l)$  und strebt nach (5.5) für  $k \rightarrow \infty$  gegen diese obere Grenze. Nach einem Satz über reelle Zahlenfolgen (falls nicht bekannt oder vergessen: Übungsaufgabe!) folgt daraus

$$\lim_{k \rightarrow \infty} kp_k = \lim_{k \rightarrow \infty} \sum_{l=1}^{\infty} P(\bar{N}_s^{(k)} \geq l) = \sum_{l=1}^{\infty} P(N_s \geq l) = EN_s = \lambda s.$$

Damit ist (5.6) gezeigt. Unser Satz folgt nun aus (5.4), (5.6) und dem Satz (5.5).  $\square$