

Frage

Die Pareto Verteilung wird benutzt, um Einkommensverteilungen zu modellieren. Die Verteilungsfunktion ist

$$F(x) = 1 - \left(\frac{k}{x}\right)^a \quad k > 0, x \geq k$$

Finden Sie den Erwartungswert und den Median der Dichte für $a > 1$. (Bei Einkommensverteilungen in der Praxis ist es üblich, dass der Erwartungswert größer als der Median ist. Warum?)

Der Gini Index mißt Einkommensungleichheit und kann u.a. als

$$C(X) = 1 - \frac{2E[X_1|X_1 < X_2]}{E[X]}$$

definiert werden, wo X_1 und X_2 u.i.v. Zufallsvariablen mit der vorgegebenen Dichte sind. Leiten Sie den Index für die Pareto Verteilung für $k=1$ her.

Anwendungen — Diskrete Wahrscheinlichkeiten

Medizinische Studien

Es werden $2N$ Personen in einer Forschungsstudie behandelt. Sie werden in zwei gleichgroßen Gruppen aufgeteilt, eine Gruppe bekommt ein aktives Medikament und die andere nur ein Placebo.

1. Unter den $2N$ Leute befinden sich $N + a$ Männer und $(N - a)$ Frauen. Was ist die Wahrscheinlichkeit, dass die zwei Gruppen denselben Anteil Männer haben, wenn das Geschlecht bei der Aufteilung nicht berücksichtigt wird?
2. Wenn die Erholungswahrscheinlichkeit für eine Person nach Einnahme des Medikaments tatsächlich p_1 beträgt und die Erholungswahrscheinlichkeit für eine Person sonst p_2 beträgt, was ist die Wahrscheinlichkeit, dass mehr Leute sich in der Placebogruppe erholen?

3. Um den Gesundheitszustand der Teilnehmer während der Studie zu überprüfen, wird jeden Tag bei jeder Person einmal ein Test durchgeführt, bis die Person den Test dreimal bestanden hat. Welche Verteilung könnte man für die Anzahl Tage nehmen, die eine Person in der Placebo Gruppe braucht, um den Test dreimal zu bestehen?

4. Es ist wohl bekannt, dass nicht alle Teilnehmer ihre Medikamente nehmen. Wenn es angenommen wird, dass die Teilnehmer in der aktiven Gruppe ihre Medikamente mit Wahrscheinlichkeit p_a nehmen, und die Teilnehmer in der Placebo Gruppe ihre Medikamente mit Wahrscheinlichkeit p_p nehmen, was ist die Wahrscheinlichkeit, dass jemand, der sich erholt hat, wirklich ein aktives Medikament eingenommen hat?

5. Es ist aus verschiedenen Gründen möglich, dass Studienteilnehmer die Studie verlassen. Welche Verteilung könnte man für die Anzahl Teilnehmer, die die Studie verlassen annehmen? Was ist die Wahrscheinlichkeit, dass mehr Leute in der Placebo Gruppe die Studie verlassen?

6. Mögliche Nebenwirkungen von Medikamenten müssen immer berichtet werden. Ob Studienteilnehmer Nebenwirkungen empfinden hängt auch von persönlichen Einstellungen ab. Wie könnte man die Anzahl berichteter Nebenwirkungen für eine Person in der Placebo Gruppe modellieren?

Autoversicherung

Versicherungsfirmen bitten ihre Kunden um viele Informationen: Geschlecht, Alter, Ehestand, Beruf, Adresse, Autotyp, Kilometer pro Jahr, Schadenfreiheitsstufe und mehr. Wieviele verschiedene Kombinationen gibt es? Die Kombinationen werden durch die Reihenfolge der Informationen definiert. Wieviele mögliche Reihenfolgen gibt es? Die Firma will Geschlecht, Schadenfreiheitsstufe, Autotyp und zwei andere Kriterien verwenden. Wieviele Möglichkeiten gibt es?

1. Wie könnte die Anzahl gemeldete Unfälle pro Jahr pro Versicherter modelliert werden?
2. Eine Firma hat N Versicherten in der Klasse A. Sei p die Unfallsrate für Versicherte dieser Klasse. Wie hängt die beobachtete Rate von N ab?
3. Die Firma benutzt ein Callcenter in Indien. Es gibt kein Warteschlangensystem für die Anrufe und die Wahrscheinlichkeit, dass es besetzt ist, beträgt r . Wie oft muss man anrufen, um durchzukommen?

4. Gegeben vier Klassen A, B, C, D. Sei die Anzahl der Versicherten in Klasse i N_i und die erwartete Unfallrate p_i , was ist die Verteilung der Anzahl Unfälle insgesamt?

5. Als Bonus am Ende des Jahres werden m Kunden ohne Unfall ausgewählt. Jeder gewinnt einen Preis. Was ist die Wahrscheinlichkeit, dass keiner aus Klasse A einen Preis gewinnt? Was ist die Wahrscheinlichkeit, dass doppelt so viele aus Klasse A einen Preis gewinnen als aus Klasse B? Gegeben, dass jemand einen Preis gewonnen hat, was ist die Wahrscheinlichkeit, dass er aus Klasse A kommt?

Wie sehen die Formeln aus, wenn man annimmt, dass die N_i 's und m groß sind?

Wer hat es geschrieben?

Die Verfasserschaft gewisser Dokumente ist unsicher. Mehrere Wissenschaftler haben Methoden vorgeschlagen, wie man unter verschiedene Kandidaten entscheiden könnte. Eins davon ist die Häufigkeiten zu untersuchen, womit Autoren gewisse einfache Wörter verwenden.

1. Gegeben sei ein Dokument von M Wörtern. Welche Verteilung könnte man für die Anzahl “also” ’s bzw. “aber” ’s nehmen?
2. Sei das apriori Glauben an den Autoren A und B durch die Wahrscheinlichkeiten p_A und $p_B = 1 - p_A$ gegeben. Aus bekannten Schriften der zwei Autoren, weißt man, dass A “also” und “aber” mit Raten λ_{A1} bzw. λ_{A2} pro 1000 Wörter verwendet.

Es gibt n_1 mal “also” im Dokument und n_2 mal “aber”. Was ist die Wahrscheinlichkeit, dass A das Dokument geschrieben hat?

3. Welche Annahmen sind hier gemacht worden?
4. Wie könnte man die Möglichkeit eines dritten unbekanntem Autors im Modell berücksichtigen?

Kleinere technische Aufgaben

- Vergleichen Sie $P(X = 0)$ für $n = 1000$ und $p = 0.001$ bei der Binomial und Poisson Verteilungen.
- Was ist der Erwartungswert einer Benfordverteilten Zufallsvariable, D ?

$$P(D = d) = \log_{10} \left(1 + \frac{1}{d} \right) \quad d = 1, \dots, 9$$

- Beweisen Sie den Satz von Bayes.
- Für welche Parameterwerte ist die Betaverteilung

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1}$$

eine Gleichverteilung?

- Was ist der Erwartungswert des Maximums von zehn unabhängigen exponentialverteilten Zufallsvariablen?

- Wie sieht die Verteilung von $Y = X^2$ aus, wenn $X \sim P(\lambda)$?

- Eine ZV X hat $E[X] = 17$ und $V[X] = 12$. Für eine Stichprobe der Größe $n = 112$ was ist $P(15 < \bar{X} < 20)$?

- Vergleichen Sie die Dichten

$$X \sim U(10, 20) \text{ und } X \sim N(17, 5)$$

im Gebiet $X \geq 17$.

- Was ist $P(X > 100)$ für
(a) X Gumbelverteilt

$$G(x) = \exp \left[-e^{-\frac{x-\xi}{\theta}} \right] \quad -\infty < x < \infty$$

- (b) X Fréchetverteilt

$$\begin{aligned} G(x) &= 0 && x < \xi \\ &= \exp \left[- \left(\frac{x - \xi}{\theta} \right)^{-\alpha} \right] && x \geq \xi \quad (\alpha > 0) \end{aligned}$$

?

Amazon ratings Sei die Verkäufe pro Woche für Buch i $X_i \sim P(\lambda_i(t))$ in der Woche t . Wenn diese Zahlen den Ranking am Samstag Abend bestimmen und $\lambda_i(t)$ und $\lambda_j(t)$ konstant sind, was ist die Wahrscheinlichkeit, dass Buch i öfters höher als Buch j an den Samstag Abenden eines Jahres gerankt wird?

Was ist die Wahrscheinlichkeit, dass mehr Kopien des Buchs j als des Buchs i im Jahr verkauft werden?

Sei

$$\lambda_i(t) = \lambda_i e^{-\nu t} \quad t = 1, \dots, 52$$

und

$$\lambda_j(t) = \lambda_j e^{-\nu t} \quad t = 1, \dots, 52$$

wie sehen die Resultate aus?

Jugendsportmodell

Sei die Dichte der Sportleistung der Kinder vom Alter a

$$f_a(x) \quad 0 \leq x$$

wo a in Tagen gemessen wird. Die besten α eines Jahrgangs (d.h. alle die in einem bestimmten Kalenderjahr geboren sind) werden für Spezialtraining ausgewählt. Wenn die Verteilung der Geburtstage gleichverteilt ist, was ist die relative Wahrscheinlichkeit, dass jemand mit einem Geburtstag im Januar ausgewählt wird, im Vergleich zu jemand mit einem Geburtstag im Dezember?

Da wir uns nur für die besten Sportler interessieren, ist die vollständige Form von $f_a(x)$ unwesentlich. Welche Formen kommen für den rechten Schwanz der Verteilung in Frage?