Universität Bremen

FB3, A. Mändle WS 2019/20, 25.11.2020

Einführung in R
7. Aufgabenblatt

Präsenzaufgabe 1

Stellen Sie die Dichtefunktionen der t-Verteilungen mit 3, 5, 10, 30 und 50 Freiheitsgraden und die Dichtefunktion einer Standard-Normalverteilung gemeinsam in einer Abbildung dar. Verwenden Sie verschiedene Farben oder/und Linientypen.

Präsenzaufgabe 2

a. Schreiben Sie eine Funktion (mit den Parametern $p, \mu_1, \mu_2, \sigma_1, \sigma_2, n$), die Zufallszahlen aus einer *gaussian mixture distribution* mit 2 Komponenten generiert.
   Die Dichtefunktion dieser Verteilung ist:

   $$f(x) = pf_1(x, \mu_1, \sigma_1) + (1-p)f_2(x, \mu_2, \sigma_2).$$

   Hinweis: Simulieren Sie `n` *Beobachtungen der gaussian mixture distribution, indem Sie zunächst* `rbinom()` *benutzen um zu bestimmen welche Anzahl* `k` *der* `n` *Simulationen aus* f_1 *kommen. Benutzen Sie dann* `rnorm()` *um* `k` *Simulationen aus* f_1 *und* `n-k` *aus* f_2 *zu erzeugen und geben Sie die Simulationsergebnisse zusammen in einer zufälligen (!) Reihenfolge als Vektor zurück.*

b. Ziehen Sie $n = 1000$ Stichproben aus

   $$X \sim 0.2 \cdot N(160, 10) + 0.8 \cdot N(180, 10)$$

   und stellen Sie die Verteilung der Daten durch eine Kerndichteschätzung dar. Fügen Sie der Grafik die theoretische Dichtefunktion hinzu.

Präsenzaufgabe 3

Angenommen man spielt 1000 Wochen lang *Lotto 6 aus 49*. Schreiben Sie eine Funktion, die berechnet, wie oft man 3 (bzw. 4, 5 oder 6) *richtige* Zahlen getippt hat.

Präsenzaufgabe 4

a. Berechnen Sie die folgenden Integrale mit der Monte-Carlo-Simulationsmethode und stellen Sie die Funktionsgraphen in den gegebenen Bereichen dar.

   $$\int_{-1}^{1} e^{x^2} dx$$

   $$\int_{0}^{\pi/8} log(1 + tan^2(x))dx$$

   $$\int_{-\pi/2}^{\pi/2} sin(2x)cos(x)dx$$

b. Stellen Sie das Konvergenzverhalten der Simulationsmethode aus a. graphisch dar: Schätzen Sie dazu das Integral für $n = 1, 2, ..., 10000$ (n= Anzahl der Simulationen) und plotten Sie die Integralschätzungen für $n = 1, 2, ..., 10000$ in einer Abbildung.
Die Funktion `integrate()` berechnet ein Integral numerisch. Verwenden Sie diese Funktion, um die o.g. Integrale zu berechnen und visualisieren Sie das Konvergenzverhalten der berechneten Werte aus der Monte Carlo-Simulation gemeinsam mit dem numerisch berechneten Wert aus der Funktion integrate().

Präsenzaufgabe 5

Wir wollen die Zahl π näherungsweise mit einer Monte Carlo-Simulationsmethode berechnen. Erzeugen Sie dazu B Zufallszahlen Z in einem Quadrat $[-1, 1] \times [-1, 1]$. Dann zählen Sie, wie viele Punkte im Kreis $K(0, 1)$ liegen. Das Verhältnis der Anzahl der Punkte im Kreis zu der Anzahl der Punkte insgesamt entspricht nährungsweise dem Verhältnis von der Fläche des Kreises zu der Fläche des Rechtecks. Dadurch lässt sich die Zahl π approximieren. Schreiben Sie eine Funktion

```
my.pi <- function(B=1000){
  x.B <- runif(B, min=?, max=?)
  y.B <- ...
  ... # Anteil der Punkte im Kreis ermitteln
  ... # Pi aus Anteil der Punkte im Kreis und Rechtecksfläche bestimmen
}
```

die dies umsetzt.

Hausaufgabe 1 (4 Punkte)

Plotten Sie die Dichtefunktionen der Gammaverteilungen mit den Parametern `shape=1:10` und `scale=1` in einer gemeinsamen Grafik. Verwenden Sie verschiedene Farben oder/und Linientypen.

Hausaufgabe 2 (6 Punkte)

Berechnen Sie die folgenden Integrale mit der Monte-Carlo-Simulationsmethode.

$$\int\limits_0^1 \log\left(1 - \tan\left(\frac{x^3}{2}\right)\right) dx$$

$$\int\limits_0^1 \int\limits_0^1 sin(x + y) dx dy$$

$$\int\limits_{-3}^3 \int\limits_{-3}^3 0.5 \cdot x^2 - 0.5 \cdot y^2 dx dy$$

Hausaufgabe 3 (10 Punkte)

Das Prinzip aus Präsenzaufgabe 5 kann auch zur Integration benutzt werden:

Gesucht ist die Fläche unter der Funktionskurve $y = f(x)$ zwischen x_1 und x_2. Wir beschreiben den Fall, dass $f(x) \geq 0$ ist. Für den Fall, dass $f(x) < 0$ im Integrationsbereich auftritt, betrachten wir die Integralfläche als Differenz der Integrale von Positiv- und Negativteil, also:

$$\int f(x)\, dx = \int \max(0, f(x))\, dx - \int |\min(0, f(x))|\, dx$$

- Man wähle ein Rechteck (i.A.: eine einfach zu berechnende Fläche), in dem die Kurve komplett eingeschlossen ist:

  $A(x_1, 0), B(x_2, 0), C(x_2, ymax), D(x_1, ymax),$

  wobei $ymax$ hier so gewählt werden muss, dass $ymax \geq f(x)$ für alle $x \in [x_1, x_2]$

- Dann generiert man N Paare von Zufallszahlen (x_i, y_i), die einer bivariaten Gleichverteilung auf diesem Rechteck folgen.

- Man bestimmt wie viele Punktepaare *Treffer* sind, d.h. $y_i \leq f(x_i)$.

- Mit dem Verhältnis der Anzahl der Treffer zu N berechnet man das Integral als

  $$\frac{\text{Trefferanzahl}}{N} \cdot \text{Rechtecksfläche}\,.$$

Verwenden Sie diese sogenannte *hit and miss* (Treffer-Nicht Treffer) Monte Carlo-Methode, um die Integrale aus Präsenzaufgabe 4 zu berechnen und vergleichen Sie die Ergebnisse.

Zusatzaufgabe 1 (5 Punkte)

Sie entscheiden sich, jede Woche einmal das Lotto-Spiel *6 aus 49* zu spielen, bis Sie zum erstem Mal mindestens 5 *Richtige* haben. Sie können nur 100 Jahre lang spielen und Sie werden jedes Mal 4 Tipps abgeben. Nehmen Sie an, dass genau 52 mal im Jahr gespielt wird.

Schreiben Sie eine Funktion, die dieses Szenario simuliert und ausgibt, wann (nach wie vielen Jahren und Wochen, wenn überhaupt) Sie das Spiel gewinnen.

Anmerkung: Sie können selbst entscheiden, wie Sie Ihre Tipps wählen, aber schlauerweise sollten die Tipps innerhalb desselben Lottospiels einer Woche nicht identisch sein.

Hinweis: machen Sie Ihr Experiment mit `set.seed()` reproduzierbar.

Abgabe der Lösungen: bis ***Montag 02.12.2019,***

maendle@uni-bremen.de