

Einführung

Das Ziel dieses Übungsblatts ist die statistische Modellierung von Protein-Seitenkettenwinkeln. Hierbei sollen mit Hilfe eines Datensatzes von Valin χ_1 Winkeln die Wahrscheinlichkeiten bestimmter Konformationen berechnet werden. Diese werden anschließend verwendet, um potentielle Seitenketten-Outlier in einer Proteinstruktur zu finden.

Chimera Attributexport

Öffnen Sie die Struktur der Protein Deglycase DJ-1 (1j42) in *Chimera*. Verschaffen Sie sich einen Überblick über das Molekül und die Qualität der Struktur. Im Folgenden liegt der Fokus ausschließlich auf Valin Aminosäuren. Alle χ_1 Winkel dieser sollen in eine Textdatei exportiert werden. Wählen Sie zunächst alle Valin Residues im Molekül aus. Starten Sie dann

Tools -> Depiction -> Render by Attribute

und von dort aus:

File -> Save Attributes

Wählen Sie unter Attribute to Save für residues den Winkel `chi1` als Attribut aus und aktivieren Sie unbedingt die Checkbox Restrict save to current selection, if any, so dass nur Daten für die ausgewählten Valin Residues gespeichert werden. Speichern Sie die Daten in einem Verzeichnis für die aktuelle Übung mit einem sinnvollen Dateinamen. Öffnen Sie die Datei mit einem Texteditor (`Kate`) und versuchen Sie das Format zu verstehen, in dem Attribute aus *Chimera* exportiert (und später wieder importiert) werden.

Rotamerdaten

Öffnen Sie die dem Übungsblatt beiliegende Datei `rota500-val.tab` mit dem Programm LibreOffice Calc. Hierbei handelt es sich um einen Ausschnitt der „*Penultimate Rotamer Library*“ (*SC Lovell, JM Word, DC Richardson (2000)*)¹. Wählen Sie beim Textimport die Sprache English (UK). Wählen Sie den Doppelpunkt als Trennsymbol für Spalten und aktivieren Sie die Erweiterte Zahlenerkennung (Detect special numbers). Betrachten Sie die Tabelle. Welche Arten von Daten liegen zusätzlich zu den Winkeln vor und wofür lassen sie sich in diesem Kontext nutzen?

Diskrete Klassen und Wahrscheinlichkeiten

Wichtig: Im Folgenden werden Sie neue Daten erzeugen. Zur Übersicht ist es sinnvoll Datenspalten, die erstellt werden sollen, mit einer entsprechenden Titelzelle zu versehen. Wenn Sie eine Rechnung durchführen, wählen Sie zunächst den Bereich aus, in dem die Ergebnisse stehen sollen, und nutzen Sie immer den Funktionsassistenten (Function Wizard - das `fx` Symbol oben links) um die gefragten Funktionen und Formeln zu erstellen. Suchen Sie hier

¹kinemage.biochem.duke.edu/databases/rotamer.php

nach der gefragten Funktion, und wählen Sie diese mit einem Doppelklick aus. Aktivieren Sie vor jeder Berechnung jeweils die Checkbox `Array` damit die Funktionsergebnisse immer zeilenweise für die ausgewählten Bereiche berechnet werden.

Die χ_1 Winkel sollen mit einer diskreten Wahrscheinlichkeitsverteilung modelliert werden. 360° werden in Intervalle eingeteilt und jedem Intervall die Wahrscheinlichkeit zugeordnet, mit der ein beliebiger Winkel in das Intervall fällt. Dies ist einfach die relative Häufigkeit der in jeweiligem Intervall beobachteten Winkel.

Entscheiden Sie sich für eine geeignete Anzahl gleich großer Intervalle, in die 360° (betrachtet von -180° bis 180°) unterteilt werden, und leiten Sie die Intervallgröße her. Erstellen Sie neben den Rotamerdaten eine Spalte, in die Sie **obere** Intervallgrenzen aufsteigend eintragen. Bei vier Intervallen und einer Intervallgröße von 90° würden entsprechend die Zahlen -90 , 0 , 90 und 180 untereinander stehen. Zur genaueren Modellierung, sollten Sie wesentlich kleinere Intervalle wählen.

Erstellen Sie direkt rechts daneben eine Spalte mit den Häufigkeiten der Winkel, die in das jeweilige Intervall fallen. Verwenden Sie dafür die Funktion `FREQUENCY()`. Der erste Parameter der Funktion (`data`) ist der Bereich mit den Winkelwerten (`chi1`) und der Zweite (`classes`) der Bereich mit den Intervallgrenzen. Sie können jeweils in das Textfeld der Parameter klicken um es zu fokussieren und anschließend mit der Maus die gewünschten Bereiche im Hauptfenster auswählen.

Erstellen Sie rechts daneben eine Spalte mit den relativen Häufigkeiten (bzw. Wahrscheinlichkeiten) der Winkel. Teilen Sie dafür den Bereich der absoluten Häufigkeiten durch seine Summe (die Sie mit der Funktion `SUM()` berechnen können). Die Formel hierfür können Sie im `Function Wizard` direkt in das `Formula` Feld eintragen. Liegen die Absoluten Häufigkeiten z.B. im Bereich `J2 : J37` so lautet die Formel: `J2:J37/SUM(J2:J37)`

Welche absoluten und relativen Häufigkeiten beobachten Sie? Reicht die Menge der Daten aus, um die relativen Häufigkeiten als Wahrscheinlichkeiten zu bezeichnen? Wie entwickeln sich die Werte über die Intervalle hinweg?

Erstellen Sie zur Übersicht ein Histogramm der relativen Häufigkeiten mit Hilfe eines Säulendiagramms. Wählen Sie hierzu die drei neu erstellten Spalten aus und klicken Sie auf `Insert -> Chart`. Wählen Sie `Column` aus und klicken Sie auf `Next>`. Aktivieren Sie im folgenden Fenster die Checkbox `First column as label` und klicken Sie wieder auf `Next>`. Entfernen Sie nun unter `Data series` die Spalte mit den absoluten Häufigkeiten (auswählen und auf `Remove` klicken) und klicken Sie auf `Finish`.

Winkel Wahrscheinlichkeitsvorhersage

Kopieren Sie die Daten (Residue Nummer und Winkel, nicht die ersten Headerzeilen) aus der mit *Chimera* erstellten Datei der Valin Winkel in die Tabelle rechts neben den Wahrscheinlichkeiten. Wählen Sie die Textimport Einstellungen wie zuvor aus, nur setzen Sie dieses Mal den Tabulator als Spaltentrennsymbol (**und nicht den Doppelpunkt**).

Zunächst sollen für die spätere Zuordnung die präzisen Werte auf ein Vielfaches Ihrer gewählten Intervallgröße aufgerundet werden. Erstellen Sie dazu rechts neben den Winkeln eine

Spalte, auf die Sie die Funktion `CEILING.MATH()` anwenden. Der erste Parameter (`Number`) ist der Bereich, der aufgerundet werden soll (die zuvor importierten Winkel) und der zweite (`Significance`) ist die Zahl auf deren Vielfaches aufgerundet wird (Ihre Intervallgröße). Vergewissern Sie sich stichprobenartig, ob das Aufrunden korrekt funktioniert (insbesondere bei negativen Zahlen).

Nun kann die Vorhersage der Winkelwahrscheinlichkeiten stattfinden. Kopieren Sie zunächst die Spalte der Residue Nummern in eine neue Spalte (dies wird später relevant). Rechts daneben sollen die Wahrscheinlichkeitswerte für die entsprechenden Residues erstellt werden. Für die Zuordnung zwischen aufgerundeten Winkeln und Wahrscheinlichkeiten der Klassen kann die Funktion `VLOOKUP()` verwendet werden. Der erste Parameter (`Search criterion`) ist dabei ein Bereich von Schlüsselwerten, für die in einem anderen Bereich zugehörige Wertepaare entsprechend rausgesucht werden sollen. Wählen Sie dafür die zuvor aufgerundeten Winkel aus. Der zweite Parameter (`array`) ist ein Bereich von mehreren Spalten, in dem sich u.A. die Wertepaare befinden, die rausgesucht werden sollen. Die erste Spalte muss dabei die Schlüsselwerte des ersten Parameters enthalten. Wählen Sie hierfür die drei am Anfang erstellten, an einander grenzenden Spalten mit den Intervallgrenzen, absoluten Häufigkeiten und Wahrscheinlichkeiten aus. Der dritte Parameter (`Index`) gibt an, in welcher Spalte des eben ausgewählten Bereichs die Daten liegen, die zu den Schlüsselwerten zugeordnet werden sollen. Tragen Sie hier die Zahl 3 ein, da in der dritten Spalte (relativ zu den Schlüsselwerten) die Wahrscheinlichkeitswerte stehen.

Es sollten nun die entsprechenden Winkelwahrscheinlichkeiten den Residue Nummern zugeordnet sein (vergewissern Sie sich wieder stichprobenartig). Kopieren Sie diese beiden Spalten und ersetzen Sie damit den Datenbereich in der aus *Chimera* exportierten Winkeldatei. Am Anfang jeder Zeile muss ein Tab-Zeichen vorliegen. Dies können Sie mit dem Texteditor *Kate* erzeugen, indem Sie zunächst in der Fußleiste unten rechts den Einrückungsmodus von Leerzeichen (`Soft Tab`) zu Tabulatoren ändern, dann den Datenbereich auswählen und mit Drücken der Tab Taste einmal einrücken. Ändern Sie in der ersten Zeile den Attributnamen von `chil` entsprechend (z.B. zu `chil_probabilities`) und speichern Sie die Datei.

Visualisierung von Wahrscheinlichkeiten

Starten Sie nun in *Chimera*

Tools -> Structure Analysis -> Define Attribute

und wählen Sie die eben gespeicherte Datei aus. Es erscheint wieder das `Render/Select by Attribute` Menü, in dem das neue Attribut der Winkelwahrscheinlichkeiten vorliegt. Färben Sie die Valin Residues nach dem Attribut.

Was beobachten Sie? Wo liegen potentielle Outlier bzw. Modellfehler? Laden Sie das besser aufgelöste Modell des gleichen Proteins `4rkw` und überlagern Sie die Strukturen mit Hilfe des `MatchMaker` Tools. Vergleichen Sie die Valin Seitenketten. Wo unterscheiden sie sich stark? Welche Werte haben die vorhergesagten Wahrscheinlichkeiten an diesen Stellen?

Vergleichen Sie wahrscheinliche mit unwahrscheinlichen Seitenketten-Konformationen und spekulieren Sie über mögliche Gründe für Ihre Beobachtungen.

Fragen

1. Es wurde ein Datensatz von χ_1 Winkeln der Valin Seitenkette verwendet. Nennen Sie einige sinnvolle Kriterien, welche die Strukturen haben sollten, aus denen ein Datensatz erstellt wird, mit welchem Winkelwahrscheinlichkeiten modelliert werden sollen.
2. Es wurde nur der Winkel χ_1 verwendet. Warum? Was ist bei der Modellierung von Wahrscheinlichkeitsverteilungen zu bedenken, wenn man weitere χ Winkel einbindet?
3. Es wurde mit einer festen Anzahl gleich großer Intervalle für die Winkelklassen gearbeitet. Gibt es dabei potentiell Probleme bei der Repräsentation der wahren Verteilung der Winkel? Was wären mögliche Verbesserungen?