Inhalt

4.1 Überblick

„Definition" bayesianischer Inferenz: Anpassen eines Wahrscheinlichkeitsmodells an eine Menge von Daten.

Ergebnis: Wahrscheinlichkeitsverteilung für die Parameter des Modells (und andere unbeobachtete Größen, zum Beispiel Vorhersagen für neue Beobachtungen).

4.1 Überblick

Idealisierter Prozess bayesianischer Datenanalyse:

1. Stelle ein *volles* Wahrscheinlichkeitsmodell oder eine gemeinsame Wahrscheinlichkeitsverteilung für alle beobachtbaren und unbeobachtbaren Größen auf. Dabei ist Wissen über das zugrundeliegende wissenschaftliche Problem und den datengenerierenden Prozess hilfreich.
2. Berechnung der Posterioriverteilung der unbeobachtbaren Größen (Parameter, missing data, ...): bedingte Wahrscheinlichkeitsverteilung der unbeobachtbaren Größen gegeben die beobachteten Daten.
3. Modelldiagnose: Fit, Sensitivität (bezüglich der Annahmen in 1.).

Ergebnis: „kohärentes System"

4.2 Exchangeability

Exchangeability („Austauschbarkeit") ist ein wichtiges Konzept für die statistische Modellbildung. Es geht auf de Finetti zurück.

Definition 4.1 (Finite Exchangeability)

Die Zufallsvariablen $X_1, \ldots, X_n$ heißen exchangeable, wenn

$$\Big(X_1, ..., X_n\Big) \text{ und } \Big(X_{\pi(1)}, ..., X_{\pi(n)}\Big) \text{ identisch verteilt sind}$$

für alle Permutationen

$$\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}.$$

Existiert eine Dichte f der gemeinsamen Verteilung, so gilt:

$$f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

4.2 Exchangeability

Definition 4.2 (Infinite Exchangeability)

Die unendliche Folge $X_1, X_2, \ldots$ *ist exchangeable, wenn jede endliche Teilfolge exchangeable ist.*

Bemerkung.
Analog zu obigen Definitionen kann auch *bedingte Exchangeability* definiert werden, etwa im Regressionsfall für $Y_1|\boldsymbol{x}_1, \ldots, Y_n|\boldsymbol{x}_n$.

Satz 4.3 (Darstellungssatz für 0-1 Zufallsvariablen)

Sei $X_1, X_2, \ldots$ eine unendliche Folge binärer Zufallsvariablen, die exchangeable sind, mit zugrundeliegendem Wahrscheinlichkeitsmaß $\mathbb{P}$.
Dann existiert eine Verteilungsfunktion Q, so dass die gemeinsame Dichte $f(x_1, \ldots, x_n)$ folgende Gestalt hat:

$$f(x_1, \ldots, x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \, dQ(\theta)$$

mit

$$Q(\theta) = \lim_{n \to \infty} \mathbb{P}\left(\frac{Y_n}{n} \leq \theta\right)$$

und

$$Y_n = \sum_{i=1}^{n} X_i$$

Interpretation:

1. Bedingt auf θ sind $X_1, X_2, \ldots$ unabhängige, identisch verteilte Bernoulli-Zufallsgrößen.
2. θ wird eine Verteilung zugeordnet.
3. Q ist der „Glaube"(„Belief") über den Grenzwert der relativen Häufigkeit der Einsen.

Konventionelle Schreibweise:

$$f(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} \theta^{x_i} (1 - \theta)^{1 - x_i}.$$

4.2 Exchangeability

Satz 4.4

Ist $X_1, X_2, \ldots$ eine (unendliche) Folge reellwertiger Zufallsvariablen, die exchangeable sind und zu denen jeweils die benötigten Dichten existieren, dann gilt

$$f(x_1, \ldots, x_n) = \int_{\Theta} \prod_{i=1}^{n} f(x_i|\theta) \, dQ(\theta).$$

Wir betrachten nun die *a posteriori prädiktive Verteilung* oder *bedingte prädiktive Verteilung* von zukünftigen (unbeobachteten) Daten $x_{m+1}, \ldots, x_n$ gegeben die beobachteten Daten $x_1, \ldots, x_m$:

$$
\begin{aligned}
f(x_{m+1},\dots,x_n|x_1,\dots,x_m) &= \frac{f(x_1,\dots,x_n)}{f(x_1,\dots,x_m)} && \text{(Satz von Bayes)}\\
&= \frac{\int\limits_\Theta \prod\limits_{i=1}^{n} f(x_i|\theta)\; dQ(\theta)}{\int\limits_\Theta \prod\limits_{i=1}^{m} f(x_i|\theta)\; dQ(\theta)} && \text{(Darstellungssatz)}\\
&= \frac{\int\limits_\Theta \prod\limits_{i=1}^{m} f(x_i|\theta) \prod\limits_{i=m+1}^{n} f(x_i|\theta)\; dQ(\theta)}{\int\limits_\Theta \prod\limits_{i=1}^{m} f(x_i|\theta)\; dQ(\theta)}\\
&= \int\limits_\Theta \prod_{i=m+1}^{n} f(x_i|\theta) \cdot \frac{\prod\limits_{i=1}^{m} f(x_i|\theta)\; dQ(\theta)}{\int\limits_\Theta \prod\limits_{i=1}^{m} f(x_i|\theta)\; dQ(\theta)}\ .
\end{aligned}
$$

4.2 Exchangeability

Dabei ist

$$\frac{\prod_{i=1}^{m} f(x_i|\theta)\; dQ(\theta)}{\int_{\Theta} \prod_{i=1}^{m} f(x_i|\theta)\; dQ(\theta)} = dQ(\theta|x_1, \ldots, x_m)$$

die Posterioriverteilung für θ gegeben Daten $x_1, \ldots, x_m$.

Hier haben wir aus „vergangenen", beobachteten Daten für zukünftige Beobachtungen gelernt. Eine Erweiterung auf andere Zufallsgrößen ist möglich:

4.2 Exchangeability

Satz 4.5 (Allgemeiner Darstellungssatz)

Sei $X_1, X_2, \ldots$ eine unendliche Folge reellwertiger Zufallsvariablen, die exchangeable sind, mit zugrundeliegendem Wahrscheinlichkeitsmaß $\mathbb{P}$.
Dann existiert ein Wahrscheinlichkeitsmaß Q über $\mathcal{F}$, dem Raum aller Verteilungsfunktionen F auf $\mathbb{R}$, so dass

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int_{\mathcal{F}} \prod_{i=1}^{n} F(x_i)\, dQ(F),$$

wobei

$$Q(F) = \lim_{n \to \infty} \mathbb{P}(F_n),$$

wobei F_n die zu $x_1, \ldots, x_n$ gehörende empirische Verteilungsfunktion bezeichnet.

Man beachte, dass obige Aussage sich (auch) auf nichtparametrische Inferenz bezieht. So steht $Q(F)$ für eine Prioriverteilung auf dem Raum aller Verteilungsfunktionen.

4.3 Bayes-Inferenz im Schnelldurchlauf

Notation:

- $\boldsymbol{X}$: beobachtete Daten
- $\widetilde{\boldsymbol{X}}$: unbeobachtete Daten
- $\boldsymbol{\theta}$: Parameter

Ziel:

- Wahrscheinlichkeitsaussagen bedingt auf beobachtete Daten
- Vorhersage / prädiktive Inferenz

Basiskomponenten in der Bayes-Inferenz:

- $p(\boldsymbol{\theta})$ Prioriverteilung
- $f(\boldsymbol{x}|\boldsymbol{\theta})$ Datenverteilung
- $f(\boldsymbol{\theta}|\boldsymbol{x})$ Posterioriverteilung
- $f(\widetilde{\boldsymbol{x}}|\boldsymbol{x})$ prädiktive Verteilung

4.3 Bayes-Inferenz im Schnelldurchlauf

Nach dem Satz von Bayes ist die gemeinsame Verteilung von $(\boldsymbol{\theta}, \boldsymbol{x})$ gleich

$$f(\boldsymbol{\theta}, \boldsymbol{x}) = f(\boldsymbol{x}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}),$$

deshalb

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{\theta}, \boldsymbol{x})}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{f(\boldsymbol{x})},$$

wobei

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{\theta \in \Theta} f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad \text{falls } \boldsymbol{\theta} \text{ diskret,} \\ f(\boldsymbol{x}) &= \int_{\Theta} f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \quad \text{falls } \boldsymbol{\theta} \text{ stetig.} \end{aligned}$$

4.3 Bayes-Inferenz im Schnelldurchlauf

Unnormalisierte Posteriori:

$$f(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

A priori prädiktive Verteilung (vor Beobachtung der Daten):

$$f(\boldsymbol{x}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{\theta}, \boldsymbol{x})\ d\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} f(\boldsymbol{x}|\boldsymbol{\theta})\ p(\boldsymbol{\theta})\ d\boldsymbol{\theta}$$

A posteriori prädiktive Verteilung (nach Beobachtung der Daten x):

$$\begin{aligned} f(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) &= \int_{\boldsymbol{\Theta}} f(\widetilde{\boldsymbol{x}}, \theta|\boldsymbol{x})\ d\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} f(\widetilde{\boldsymbol{x}}|\boldsymbol{\theta}, \boldsymbol{x})\ f(\boldsymbol{\theta}|\boldsymbol{x})\ d\boldsymbol{\theta} \\ &= \int_{\boldsymbol{\Theta}} f(\widetilde{\boldsymbol{x}}|\boldsymbol{\theta})\ f(\boldsymbol{\theta}|\boldsymbol{x})\ d\boldsymbol{\theta}, \end{aligned}$$

da $\widetilde{\boldsymbol{x}}$ bedingt unabhängig von $\boldsymbol{x}$ gegeben $\boldsymbol{\theta}$ ist.

4.3 Bayes-Inferenz im Schnelldurchlauf

Likelihood und Odds Ratios
Die Likelihoodfunktion ist $f(x|\theta)$ als Funktion von $\boldsymbol{\theta}$ nach Beobachtung von *bx*. Die Daten beeinflussen die Posteriori-Inferenz also nur über die Likelihood.
Die Posteriori-Odds von $\boldsymbol{\theta}_1$ verglichen mit $\boldsymbol{\theta}_2$ sind

$$\frac{f_{\theta|x}(\boldsymbol{\theta}_1|\boldsymbol{x})}{f_{\theta|x}(\boldsymbol{\theta}_2|\boldsymbol{x})} = \frac{\frac{f(\boldsymbol{x}|\boldsymbol{\theta}_1)f(\boldsymbol{\theta}_1)}{f(\boldsymbol{x})}}{\frac{f(\boldsymbol{x}|\boldsymbol{\theta}_2)f(\boldsymbol{\theta}_2)}{f(\boldsymbol{x})}} = \frac{f(\boldsymbol{x}|\boldsymbol{\theta}_1)}{f(\boldsymbol{x}|\boldsymbol{\theta}_2)} \cdot \frac{f(\boldsymbol{\theta}_1)}{f(\boldsymbol{\theta}_2)},$$

es gilt also

$$\text{Posteriori-Odds} = \text{Priori-Odds} \times \text{Likelihoodquotient}.$$

4.4 Wiederholung: Modelle mit einem Parameter

- Gemeint ist: θ ist skalar.
- Prioriverteilung kann mehr als einen Parameter haben.
- Hier funktionieren folgende Konzepte gut:
  - **Konjugierte Prioriverteilungen**
  - **Referenzprioris/Referenzanalyse:**

4.4 Wiederholung: Modelle mit einem Parameter

- **Konjugierte Prioriverteilungen**, zum Beispiel bei (einparametrischen) Exponentialfamilien.

  Vorteil: Analytische Berechenbarkeit, keine Simulation nötig.

  Nachteil: Für komplexe Modelle meist nicht verfügbar, deshalb eher als Baustein in komplizierteren Modellen verwendet.

4.4 Wiederholung: Modelle mit einem Parameter

- **Referenzprioris/Referenzanalyse:**
  - *Idee:* Priori so wählen, dass die Daten auch im Fall geringen Stichprobenumfangs die Posterioriverteilung dominieren („let the data speak for themselves"). Dies benötigt entscheidungs- und informationstheoretische Grundlagen.
    Suche nach nicht-informativen Prioriverteilungen: Im skalaren Fall zum Beispiel

$$0 < \theta < 1 \;\rightarrow\; \psi = \log\left(\frac{\theta}{1-\theta}\right).$$

  - *Jeffreys' Priori*

$$p(\theta) \propto \sqrt{\mathcal{I}(\theta)}$$

    ist invariant gegenüber bijektiven Transformationen von θ.

4.4 Wiederholung: Modelle mit einem Parameter

Beispiel 4.1 (Binomial- und Negative Binomialverteilung)

1. *Binomialverteilung:* Die Likelihood lautet

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}.$$

Als Referenzpriori kann Jeffreys' Priori, Beta $\left(\frac{1}{2}, \frac{1}{2}\right)$, verwendet werden:

$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}.$$

Sei $y = \sum_{i=1}^{n} x_i$. Dann ist die Referenzposteriori:

$$\begin{aligned} f(\theta|\boldsymbol{x}) &\propto f(\boldsymbol{x}|\theta)p(\theta) \\ &\propto \theta^{y}(1-\theta)^{n-y}\theta^{-1/2}(1-\theta)^{-1/2} \\ &= \theta^{y-1/2}(1-\theta)^{n-y-1/2}. \end{aligned}$$

1. *fortgeführt*
   Dies entspricht dem Kern der Dichte einer
   Beta $\left(\frac{1}{2}+y, \frac{1}{2}+n-y\right)$–Verteilung. $f(\theta|\boldsymbol{x})$ ist auch für die Extremfälle $y=0$ oder $y=n$ noch „proper".

   Verwendet man dagegen Haldane's Priori

   $$p(\theta) \propto \theta^{-1}(1-\theta)^{-1},$$

   die eine uneigentliche Priori 'Beta$(0,0)$' darstellt, ist die Posteriori Beta$(y, n-y)$ für die Extreme $y=0$ oder $y=n$ nicht proper.

2. *Negative Binomialverteilung:* Sei X die Anzahl an Versuchen bis $y \geq 1$ Erfolge eintreten. Dann lautet die Likelihood

$$f(x|\theta) \propto \binom{x-1}{y-1} \theta^y (1-\theta)^{x-y} \quad \text{für } x \geq y.$$

Die Referenzpriori ist durch Jeffreys' Priori für die geometrische Verteilung gegeben (das entspricht $y = 1$):

$$p(\theta) \propto \theta^{-1}(1-\theta)^{-1/2},$$

woraus die Referenzposteriori

$$f(\theta|x) \propto \theta^{y-1}(1-\theta)^{x-y-1/2},$$

also eine Beta$(y, x-y+1/2)$–Verteilung, resultiert.
Da $y \geq 1$ und $x \geq y$, ist auch diese a posteriori stets proper.

Bemerkung.
Konzepte für eindimensionale Modelle sind im mehrdimensionalen Fall im Allgemeinen schwierig umzusetzen bzw. umstritten (zum Beispiel Verwendung von Referenzprioris). Man geht daher oft zu sogenannten hierarchischen Modellen über: Füge zusätzliche Stufen in das Modell ein mit dem Ziel, die Posteriori-Analyse stärker von Priori-Annahmen zu entkoppeln.

4.5 Mehr-Parameter-Modelle

4.5.1 Normalverteilung

Wir betrachten in diesem Abschnitt Daten $x_1, \ldots, x_n | \mu, \sigma^2 \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ mit μ, σ^2 unbekannt.

(i) *Gemeinsame Posterioriverteilung von $\mu, \sigma^2 | \boldsymbol{x}$:*

$$f(\mu, \sigma^2 | \boldsymbol{x}) \propto f(\boldsymbol{x} | \mu, \sigma^2) \cdot p(\mu, \sigma^2)$$

(ii) *Bedingte Posterioriverteilungen von $\mu | \sigma^2, \boldsymbol{x}$ bzw. $\sigma^2 | \mu, \boldsymbol{x}$:*

$$f(\mu | \sigma^2, \boldsymbol{x}) \quad \text{bzw.} \quad f(\sigma^2 | \mu, \boldsymbol{x})$$

(iii) *Marginale Posterioriverteilung von $\mu | \boldsymbol{x}$:*

$$f(\mu | \boldsymbol{x}) = \int f(\mu, \sigma^2 | \boldsymbol{x})\, d\sigma^2 = \int f(\mu | \sigma^2, \boldsymbol{x}) f(\sigma^2 | \boldsymbol{x})\, d\sigma^2$$

I. Nichtinformative Prioriverteilung
Ist nur einer der beiden Parameter unbekannt, so wählt man oft folgende **Prioriverteilungen** (Jeffreys' Prioris):

$$\begin{aligned} \sigma^2 \text{ bekannt:} \quad & p(\mu) \propto \text{ const}, \\ \mu \text{ bekannt:} \quad & p(\sigma^2) \propto (\sigma^2)^{-1}. \end{aligned}$$

Eine Möglichkeit, daraus eine mehrdimensionale Priori zu konstruieren, ist:

$$p(\mu, \sigma^2) = p(\mu) \cdot p(\sigma^2) \propto (\sigma^2)^{-1},$$

d.h. wir nehmen unabhängige Prioris für μ und σ^2 an.

Die **gemeinsame Posterioriverteilung** $f(\mu, \sigma^2|\boldsymbol{x})$ lautet dann:

$$
\begin{aligned}
f(\mu, \sigma^2|\boldsymbol{x}) &\propto \text{Likelihood} \times \text{Priori} \\
&= \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \sigma^{-1} \exp\left(-\frac{1}{2\sigma^2} bx_i - \mu)^2\right)\right] \cdot (\sigma^2)^{-1} \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left((n-1)s^2 + n(\overline{x} - \mu)^2\right)\right)
\end{aligned}
$$

mit $s^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2/(n-1)$.

Die **bedingte Posteriori** von μ, $f(\mu|\sigma^2, \boldsymbol{x})$, kann auf den Fall mit bekannter Varianz σ^2 zurückgeführt werden. Aus Statistik IV ist bekannt, dass $f(\mu|\sigma^2, \boldsymbol{x}) \sim N(\overline{x}, \sigma^2/n)$.
Für die **marginale Posteriori** $f(\sigma^2|\boldsymbol{x})$ hat man

$$
\begin{aligned}
f(\sigma^2|\boldsymbol{x}) &= \int f(\mu, \sigma^2|\boldsymbol{x})\, d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\overline{x}-\mu)^2\right)\right)\, d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \int \exp\left(-\frac{1}{2\sigma^2} n(\overline{x}-\mu)^2\right) d\mu
\end{aligned}
$$

Es gilt

$$\int \exp\left(-\frac{1}{2\sigma^2} n(\overline{x}-\mu)^2\right) \, d\mu = \sqrt{2\pi \, \frac{\sigma^2}{n}}$$

und damit

$$\begin{aligned}
f(\sigma^2|\boldsymbol{x}) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi \, \frac{\sigma^2}{n}} \\
&\propto (\sigma^2)^{\frac{-n-2}{2}} (\sigma^2)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \\
&= (\sigma^2)^{-\frac{n+1}{2}} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right).
\end{aligned}$$

Der Kern dieser Dichte gehört zur inversen Gamma-Verteilung mit den Parametern $(n-1)/2$ und $(n-1)s^2/2$.
Wegen

$$f(\mu, \sigma^2 | x_1, \ldots, x_n) = f(\mu | \sigma^2, x_1, \ldots, x_n) \cdot f(\sigma^2 | x_1, \ldots, x_n)$$

kann die gemeinsame Posterioriverteilung von $\mu, \sigma^2 | x_1, \ldots, x_n$ nun wie folgt simuliert werden:

Algorithmus 2 : Direkte Simulation der gemeinsamen Posteriori-verteilung bei nichtinformativer Priori

Wiederhole für $s = 1, \ldots, S$:

Schritt 1: Ziehe $(\sigma^2)^{(s)}$ aus $\mathrm{IG}\left(\frac{n-1}{2}, \frac{n-1}{2}s^2\right)$.

Schritt 2: Ziehe $(\mu)^{(s)}$ aus $N(\overline{x}, \frac{1}{n}(\sigma^2)^{(s)})$.

Man erhält Paare $\left[(\mu^{(1)}, (\sigma^2)^{(1)}), \ldots, (\mu^{(S)}, (\sigma^2)^{(S)})\right]$.

$\boldsymbol{\sigma}^2$ als Nuisance-Parameter

Interessiert nur μ, so gibt es (mindestens) zwei Möglichkeiten zur Simulation:

1. Simuliere die gemeinsame Posteriori $f(\mu, \sigma^2|\boldsymbol{x})$ gemäß obigem Algorithmus und betrachte nur die Ziehungen $\mu^{(1)}, \ldots, \mu^{(S)}$.

2. Berechne direkt die marginale Posteriori $f(\mu|\boldsymbol{x})$:

$$f(\mu|\boldsymbol{x}) = \int_0^\infty f(\mu, \sigma^2|\boldsymbol{x})\, d\sigma^2.$$

Zu 2. Führt man die Substitution $z = A/(2\sigma^2)$ mit $A = (n-1)s^2 + n(\mu - \overline{x})^2$ durch, so erhält man wegen

$$\sigma^2 = \frac{1}{2}Az^{-1} \quad \text{und} \quad d\sigma^2 = -2A^{-1}\sigma^4 dz = -\frac{1}{2}Az^{-2}dz$$

für $f(\mu|\boldsymbol{x})$

$$\begin{aligned}
\int_0^\infty f(\mu, \sigma^2|\boldsymbol{x})\, d\sigma^2 &\propto \int_0^\infty A^{-\frac{n+2}{2}} z^{\frac{n+2}{2}} \exp(-z)\, A\, z^{-2} dz \\
&= \int_0^\infty A^{-\frac{n}{2}} z^{\frac{n-2}{2}} \exp(-z)\, dz \\
&= A^{-\frac{n}{2}} \int_0^\infty z^{\frac{n-2}{2}} \exp(-z)\, dz.
\end{aligned}$$

Zu 2. Allgemein gilt für $a > 0$ und $m > -1$:

$$\int_0^\infty z^m \exp(-az)\, dz = \frac{\Gamma(m+1)}{a^{m+1}}.$$

Daraus folgt, dass das Integral konstant bezüglich μ ist und somit

$$\begin{aligned} f(\mu|\boldsymbol{x}) &\propto A^{-\frac{n}{2}} \\ &= \left[(n-1)s^2 + n(\mu - \overline{x})^2\right]^{-\frac{n}{2}} \\ &= \left[1 + \frac{(\mu - \overline{x})^2}{(n-1)s^2/n}\right]^{-\frac{n}{2}} \\ &= \left[1 + \frac{1}{n-1}\left(\frac{\mu - \overline{x}}{\frac{s}{\sqrt{n}}}\right)^2\right]^{-\frac{n}{2}} \end{aligned}$$

...

Zu 2. ...

was der Kern einer skalierten nichtzentralen t-Verteilung mit Skalenparameter $m = s/\sqrt{n}$, Lokationsparameter $l = \overline{x}$ und $\nu = n - 1$ Freiheitsgraden ist.

Allgemein hat der Kern der Dichte einer solchen allgemeinen t-Verteilung die Gestalt

$$f(\theta) \propto \left[1 + \frac{1}{\nu}\left(\frac{\theta - l}{m}\right)^2\right]^{-\frac{\nu+1}{2}} .$$

Bemerkung

1. Statt σ^2 lässt sich auch die sogenannte *Präzision* $\kappa = (\sigma^2)^{-1}$ verwenden. Bei Verwendung von $p(\mu, \kappa) \propto (\kappa)^{-1}$ folgt, dass $\kappa | x \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{n-1}{2} s^2\right)$.
2. Statt inverser Gammaverteilung wird häufig der Spezialfall einer sogenannten skalierten inversen χ^2-Verteilung inv-χ^2 verwendet (siehe unten).

II. Konjugierte Prioriverteilung

Verwende gemäß Bemerkung 2 die skalierte inverse $\chi^2(\nu_0, \sigma_0^2)$-Verteilung als Priori.

Vorteil: Bessere Interpretation (das werden wir allerdings erst dann verstehen, wenn wir informative Prioriverteilungen in Form von (inversen) Gammaverteilungen betrachten).

Nachteil: Diese Vorgehensweise ist „non-standard".

Zufallszahlen aus einer skalierten inversen χ^2 -Verteilung kann man wie folgt simulieren:

Algorithmus 3 : Simulation von $\boldsymbol{\theta \sim}$ inv-$\boldsymbol{\chi^2(\nu_0, \sigma_0^2)}$

---

1. Ziehe $X^* \sim \chi^2(\nu_0)$.
2. Setze $\theta = \frac{\nu_0 \sigma_0^2}{X^*}$.

Es gilt:

$$\text{inv-}\chi^2(\nu_0, \sigma_0^2) \; = \; \mathsf{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_0^2\right).$$

Dies lässt sich mit dem Transformationssatz für Dichten verifizieren: Definiere $\alpha = \nu_0/2$ und $\beta = 1/2$, so dass $X^* \sim \text{Gamma}(\alpha, \beta)$.

Die Umkehrfunktion der Transformation in Schritt 2 lautet

$$X^* = g^{-1}(\theta) = \frac{\nu_0 \sigma_0^2}{\theta}$$

und die zugehörige Ableitung nach θ

$$(g^{-1})'(\theta) = -\frac{\nu_0 \sigma_0^2}{\theta^2}.$$

Man erhält somit:

$$f(\theta) = f_{X^*}(g^{-1}(\theta)) \cdot |(g^{-1})'(\theta)| = \frac{(\beta \nu_0 \sigma_0^2)^\alpha}{\Gamma(\alpha)} \, \theta^{-\alpha-1} \exp\left(-\beta \nu_0 \sigma_0^2 / \theta\right).$$

Dies ist die Dichte einer inversen Gammaverteilung mit Parametern $(\alpha, \beta\nu_0\sigma_0^2)$, welche gerade der gewünschten inversen χ^2-Verteilung entspricht. Eine mögliche Parametrisierung ist nun

$$p(\mu, \sigma^2) = p(\mu|\sigma^2) \cdot p(\sigma^2)$$

mit

$$\mu|\sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \quad \text{und} \quad \sigma^2 \sim \text{inv-}\chi^2(\nu_0, \sigma_0^2)$$

Man schreibt hierfür kurz: N-inv-$\chi^2\left(\mu_0, \frac{\sigma_0^2}{\kappa_0}; \nu_0, \sigma_0^2\right)$. Die Prioris sind nunmehr voneinander abhängig.

Sei nun also a priori $(\mu, \sigma^2) \sim$ N-inv-$\chi^2\left(\mu_0, \frac{\sigma_0^2}{\kappa_0}; \nu_0, \sigma_0^2\right)$.

Die **Prioridichte** lautet dann

$$
\begin{aligned}
p(\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2\kappa_0^{-1}}} \exp\left(-\frac{1}{2\frac{\sigma^2}{\kappa_0}}(\mu-\mu_0)^2\right) \\
&\times \frac{\left(\frac{1}{2}\nu_0\sigma_0^2\right)^{\nu_0/2}}{\Gamma(\nu_0/2)} (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(-\frac{1}{2}\frac{\nu_0\sigma_0^2}{\sigma^2}\right) \\
&\propto (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2 + \kappa_0(\mu-\mu_0)^2\right]\right).
\end{aligned}
$$

Die **gemeinsame Posteriori** bei gegebenen Daten $\boldsymbol{x} = (x_1, \ldots, x_n)$ aus $N(\mu, \sigma^2)$ ergibt sich zu:

$$\begin{aligned} f(\mu, \sigma^2 | \boldsymbol{x}) \quad &\propto \quad (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(-\frac{1}{2\sigma^2} \left[\nu_0 \sigma_0^2 + \kappa_0 (\mu - \mu_0)^2 \right] \right) \\ &\quad \times (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \left((n-1)s^2 + n(\overline{x} - \mu)^2 \right) \right). \end{aligned}$$

Man kann zeigen (vgl. Übung), dass die Posteriori wieder N-inv-χ^2-verteilt ist mit Parametern

$$
\begin{aligned}
\mu_n &= \left(\frac{\kappa_0}{\kappa_0 + n}\right)\mu_0 + \left(\frac{n}{\kappa_0 + n}\right)\overline{x}, \\
\kappa_n &= \kappa_0 + n, \\
\nu_n &= \nu_0 + n, \\
\nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\overline{x} - \mu_0)^2.
\end{aligned}
$$

Die Interpretation der Parameter ist wie folgt:

- μ_n ist gewichteter Mittelwert aus Stichprobenmittel und Priori-Erwartungswert. In den Grenzfällen $\kappa_0 \to \infty$ ist $\mu_n = \mu_0$ bzw. für $n \to \infty$ ist $\mu_n = \overline{x}$.
- ν_n sind die Posteriori-Freiheitsgrade als Summe von Priori-Freiheitsgraden und Stichprobenumfang.
- Die Posteriori-Quadratsumme $\nu_n \sigma_n^2$ lässt sich partitionieren in die Priori-Quadratsumme $\nu_0 \sigma_0^2$, die Quadratsumme $(n-1)s^2$ der Stichprobe und einen Term, der die Unsicherheit, die durch die Differenz zwischen Stichprobenmittel und Priori-Erwartungswert ensteht, quantifiziert.

Die **bedingte Posteriori** von $\mu|\sigma^2, \boldsymbol{x}$ ist

$$\begin{aligned}
\mu|\sigma^2, \boldsymbol{x} &\sim N\left(\mu_n, \frac{\sigma^2}{\kappa_n}\right) \\
&\hat{=} N\left(\frac{\kappa_0}{\kappa_0 + n}\,\mu_0 + \frac{n}{\kappa_0 + n}\,\bar{x}, \frac{\sigma^2}{\kappa_0 + n}\right) \\
&\hat{=} N\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}\right).
\end{aligned}$$

Die Gewichte κ_0/σ^2 und n/σ^2 entsprechen der Priori- bzw. Datenpräzision. Die **marginale Posteriori** von $\sigma^2|\boldsymbol{x}$ ist

$$\sigma^2|\boldsymbol{x} \sim \text{inv-}\chi^2(\nu_n, \sigma_n^2).$$

Dies ermöglicht die Simulation der gemeinsamen Posteriori Verteilung:

Algorithmus 4 : Direkte Simulation der gemeinsamen Posterioriverteilung bei konjugierter Priori

---

Schritt 1: Ziehe $(\sigma^2)^*$ aus inv-$\chi^2(\nu_n, \sigma_n^2)$.

Schritt 2: Ziehe μ^* aus $N\left(\mu_n, \frac{(\sigma^2)^*}{\kappa_n}\right)$.

Die **marginale Posteriori** von $\mu|x$ lautet

$$f(\mu|\boldsymbol{x}) \propto \left[1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n \sigma_n^2}\right]^{-(\nu_n+1)/2} .$$

Dies entspricht einer t-Verteilung mit ν_n Freiheitsgraden, Lokationsparameter μ_n und Skalenparameter σ_n^2/κ_n.

III. Semi-konjugierte Prioriverteilung
Die Parameter μ und σ^2 sollen nun a priori unabhängig sein. Wir wählen deshalb **a priori**

$$\mu|\sigma^2 \;\sim\; N(\mu_0, \tau_0^2) \quad \text{und} \quad \sigma^2 \;\sim\; \text{inv-}\chi^2(\nu_0, \sigma_0^2).$$

Der einzige Unterschied zum Fall der konjugierten Priori ist also, dass wir τ_0^2 statt σ_0^2/κ_0 verwenden und so die Prioris entkoppeln. Es folgt:

$$p(\mu, \sigma^2) = p(\mu|\sigma^2) \cdot p(\sigma^2) \overset{\text{Unabhängigkeit}}{=} p(\mu) \cdot p(\sigma^2).$$

Die resultierende gemeinsame Posteriori hat allerdings keine Form, die einer bekannten Verteilung zugeordnet werden kann.
Allerdings ist $f(\mu|\sigma^2, \boldsymbol{x})$ explizit berechenbar und $f(\sigma^2|\boldsymbol{x})$ einfach zu simulieren.
Die **bedingte Posteriori** ist $\mu|\sigma^2, \boldsymbol{x} \sim N(\mu_n, \tau_n^2)$ mit

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\overline{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad \text{und} \quad \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}.$$

Zur Herleitung der Posteriori $f(\sigma^2|\boldsymbol{x})$ benutzt man, dass

$$f(\sigma^2|\boldsymbol{x}) = \frac{f(\mu, \sigma^2|\boldsymbol{x})}{f(\mu|\sigma^2, \boldsymbol{x})}$$

und (salopp), dass

$$f(\mu, \sigma^2|\boldsymbol{x}) \propto N(\mu|\mu_0, \tau_0^2) \times \text{inv-}\chi^2(\sigma^2|\nu_0, \sigma_0^2) \times \prod_{i=1}^{n} N(x_i|\mu, \sigma^2).$$

Die **marginale Posteriori** hat dann die Struktur

$$f(\sigma^2|\boldsymbol{x}) \propto \frac{N(\mu|\mu_0,\tau_0^2)\cdot \text{inv-}\chi^2(\sigma^2|\nu_0,\sigma_0^2)\cdot \prod_{i=1}^n N(x_i|\mu,\sigma^2)}{N(\mu|\mu_n,\tau_n^2)}.$$

Da $f(\sigma^2|\boldsymbol{x})$ nicht von μ abhängen kann, können die entsprechenden Terme ignoriert werden, die nur von μ abhängen. Man beachte jedoch, dass der Nenner über die Parameter μ_n, τ_n^2 noch von σ^2 abhängt. Setzen wir $\mu = \mu_n$, so erhalten wir

$$f(\sigma^2|\boldsymbol{x}) \propto \tau_n \exp\left(-\frac{1}{2\tau_0^2}(\mu_n-\mu_0)^2\right)\cdot \text{inv-}\chi^2(\sigma^2|\nu_0,\sigma_0^2)$$
$$\cdot \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_n)^2\right).$$

Diese Verteilung lässt sich beispielsweise über einen empirischen CDF-Sampler simulieren. Voraussetzung hierfür ist, dass wir die Dichte einer (univariaten) Zufallsvariable bis auf eine Proportionalitätskonstante c kennen.

Algorithmus 5 : Empirischer CDF-Sampler

- Diskretisiere den Träger der zu simulierenden Verteilung in eine Menge von N Punkten $x_1 \leq \ldots \leq x_N$.
- Evaluiere die bis auf Proportionalität bekannte Dichte an $x_1 \leq \ldots \leq x_N$, um Werte $f_1, \ldots, f_N$ zu erhalten.
- Schätze die Proportionalitätskonstante c über $c = f_1 + \ldots + f_N$.
- Ziehe Zufallszahlen aus $x_1 \leq \ldots \leq x_N$ gemäß den Wahrscheinlichkeiten $f_1/c, \ldots, f_N/c$.

Dies führt zu folgendem Algorithmus zur Simulation aus $f(\mu, \sigma^2|\boldsymbol{x})$:

1. Ziehe $(\sigma^2)^*$ aus der marginalen (approximativen) Posteriori gemäß CDF-Sampler.
2. Ziehe μ aus $f(\mu|(\sigma^2)^*, \boldsymbol{x})$.

Abschließend betrachten wir noch die **prädiktive Posterioriverteilung** für zukünftige Beobachtungen $\widetilde{\boldsymbol{x}}$ gegeben Daten $\boldsymbol{x}$. Diese lautet

$$
\begin{aligned}
f(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) &= \int\int f(\widetilde{\boldsymbol{x}}|\mu, \sigma^2, \boldsymbol{x}) f(\mu, \sigma^2|\boldsymbol{x}) \; d\mu \; d\sigma^2 \\
&= \int\int f(\widetilde{\boldsymbol{x}}|\mu, \sigma^2) f(\mu, \sigma^2|\boldsymbol{x}) \; d\mu \; d\sigma^2.
\end{aligned}
$$

Im *Dirichlet-Multinomial-Modell* wird für die Daten $y_1, \ldots, y_n$ eine *Multinomialverteilung* angenommen, also die Verallgemeinerung der Binomialverteilung auf mehr als zwei mögliche Ereignisse bei festem Stichprobenumfang n. Beispielsweise könnte eine fest vorgegebene Anzahl an Personen nach ihrer Parteipräferenz befragt werden.

Eine multinomialverteilte Zufallsvariable Y kann k mögliche Ausprägungen annehmen (zum Beispiel CDU/CSU, SPD, FDP, Grüne, Linke, andere).
Die Zufallsvariable X_j, $1 \leq j \leq k$, bezeichnet die Anzahl der j-ten Ausprägung in der Stichprobe; es gilt $\sum_{j=1}^{k} X_j = n$. Der Parameter $\theta_j = \mathbb{P}(Y = j) \in [0, 1]$ für $Y \in \{1, \ldots, k\}$ mit $\sum_{j=1}^{k} \theta_j = 1$ bezeichnet die Wahrscheinlichkeit für die Ausprägung j.

Die Likelihood von $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$ bei Beobachtungen $\boldsymbol{x} = (x_1, \ldots, x_k)^\top$ lautet

$$L(\boldsymbol{\theta}) = f(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{j=1}^{k} \theta_j^{x_j}.$$

Wegen der Restriktion $\sum_{j=1}^{k} \theta_j = 1$ liegen faktisch nur $k-1$ Parameter vor, denn der k-te Parameter lässt sich deterministisch durch $\theta_k = 1 - \theta_1 - \ldots \theta_{k-1}$ berechnen. Die Likelihood lässt sich daher auch in der Form

$$L(\boldsymbol{\theta}) \propto \left(\prod_{j=1}^{k-1} \theta_j^{x_j} \right) (1 - \theta_1 - \ldots - \theta_{k-1})^{x_k}$$

schreiben.

Die zur Multinomialverteilung konjugierte Verteilung ist die sogenannte *Dirichletverteilung*, eine Verallgemeinerung der Beta-Verteilung, geschrieben

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k) = D(\boldsymbol{\alpha}),$$

mit Dichtefunktion

$$p(\boldsymbol{\theta}) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \cdot \ldots \cdot \Gamma(\alpha_k)} \, \theta_1^{\alpha_1 - 1} \cdot \ldots \cdot \theta_k^{\alpha_k - 1},$$

wobei $\alpha_j > 0$ für alle $j = 1, \ldots, k$ und wieder $\theta_j \in [0, 1]$ mit $\sum_{j=1}^k \theta_j = 1$. Die Dirichlet-Verteilung spezifiziert also eine Verteilung auf einem $(k-1)$-dimensionalen offenen Simplex.

Eigenschaften der Dirichletverteilung:

Definiere $\alpha_0 = \sum_{j=1}^{k} \alpha_j$.

- Momente:

$$\mathbb{E}(\theta_j) = \frac{\alpha_j}{\alpha_0},$$
$$\mathrm{Var}(\theta_j) = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)},$$
$$\mathrm{Cov}(\theta_i, \theta_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)} \; ,$$

wobei die Restriktion $\sum_{j=1}^{k} \theta_j = 1$ die negative Korrelation impliziert.

- Modus:

$$\text{Modus}(\boldsymbol{\theta})_j = \frac{\alpha_j - 1}{\alpha_0 - k}$$

ist die j-te Komponente des k-dimensionalen Modus.

- Jede Randverteilung ist wieder eine Dirichletverteilung, zum Beispiel

$$(\theta_i, \theta_j, 1 - \theta_i - \theta_j) \;\sim\; \text{Dirichlet}(\alpha_i, \alpha_j, \alpha_0 - \alpha_i - \alpha_j).$$

Insbesondere ist

$$\theta_j \;\sim\; \text{Beta}(\alpha_j, \alpha_0 - \alpha_j).$$

- Die bedingten Verteilungen sind ebenfalls Dirichletverteilt. Setzt man

$$\theta'_i = \frac{\theta_m}{1 - \sum_{r=1}^{m-1} \theta_r} \quad , \; m \leq \; i \leq k,$$

gegeben die Realisationen $\theta_1, \ldots, \theta_{m-1}$, so ist

$$(\theta'_m, \ldots, \theta'_k)^\top \sim \text{Dirichlet}(\alpha_m, \ldots, \alpha_k).$$

Algorithmus 6 : Simulation aus der Dirichletverteilung

- Simulation 1:
  1. Ziehe $Z_1, Z_2, \ldots, Z_k$ aus (unabhängigen) Gammaverteilungen mit Parametern $(\alpha_1, 1), \ldots, (\alpha_k, 1)$.
  2. Setze
  $$\theta_j = \frac{Z_j}{\sum_{i=1}^{k} Z_i} \; .$$
- Simulation 2 („Stick Breaking Prior"):
  1. Ziehe $\theta_1 \sim \text{Beta}(\alpha_1, \alpha_0 - \alpha_1)$.
  2. Für $j = 2, \ldots, k-1$:
     (i) Ziehe $Z_j \sim \text{Beta}(\alpha_j, \sum_{i=j+1}^{k} \alpha_i)$.
     (ii) Setze $\theta_j = \left(1 - \sum_{i=1}^{j-1} \theta_i\right) Z_j$.
  3. Setze $\theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$.

Für $\boldsymbol{x}|\boldsymbol{\theta} \sim \text{Multinomial}(n; \theta_1, \ldots, \theta_k)$ und $\boldsymbol{\theta}|\boldsymbol{\alpha} \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$ lautet die **Posterioriverteilung** von $\boldsymbol{\theta}$:

$$
\begin{aligned}
f(\boldsymbol{\theta}|\boldsymbol{x}) &\propto L(\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}|\boldsymbol{\alpha}) \\
&\propto \prod_{j=1}^{k} \theta_j^{x_j} \cdot \prod_{j=1}^{k} \theta_j^{\alpha_j - 1} \\
&= \prod_{j=1}^{k} \theta_j^{x_j + \alpha_j - 1},
\end{aligned}
$$

d.h. die Posteriori ist Dirichlet$(x_1 + \alpha_1, \ldots, x_k + \alpha_k)$-verteilt.

Interpretation der Posteriori

Der Posteriori-Erwartungswert

$$\mathbb{E}[\theta_j|\boldsymbol{x}] = \frac{x_j + \alpha_j}{\sum_{i=1}^k x_i + \sum_{i=1}^k \alpha_i} = \frac{x_j + \alpha_j}{n + \alpha_0}$$

lässt sich umschreiben zu

$$\mathbb{E}[\theta_j|\boldsymbol{x}] = \frac{\alpha_0}{\alpha_0 + n} \cdot \underbrace{\frac{\alpha_j}{\alpha_0}}_{\substack{\text{Priori-}\\ \text{Erwartungswert}}} + \frac{n}{\alpha_0 + n} \cdot \underbrace{\frac{x_j}{n}}_{\text{MLE}} .$$

Der Parameter α_0 lässt sich als „a priori Anzahl Beobachtungen” und α_j als „a priori Erfolge” für Kategorie j interpretieren.

Bemerkung.

1. Die Wahl $\alpha_1 = \ldots = \alpha_k = 0$ entspricht einer Gleichverteilung für $\{\log \theta_j\}_{j=1}^k$. In diesem Fall ist die Posteriori nur dann proper, wenn $x_j \geq 1,\ j = 1, \ldots, p$.
2. Die Wahl $\alpha_1 = \ldots = \alpha_k = 1/2$ entspricht Jeffreys' Priori.
3. Die Wahl $\alpha_1 = \ldots = \alpha_k = 1$ entspricht einer Priori-Gleichverteilung auf dem Simplex.

Bemerkung.
Die Dirichlet-Verteilung eignet sich auch als Priori bei der Analyse von Kontigenztafeln mit multinomialem Erhebungsschema:

x_1	x_2
x_3	x_4

($n = x_1 + x_2 + x_3 + x_4$). Erweiterungen auf restringierte Multinomialverteilungen sind möglich (loglineare Modelle).
...

...

Ad-hoc Prozedur: Addiere $1/2$ zu jedem Eintrag der Kontingenztabelle und berechne dann den Maximum-Likelihood-Schätzer; das entspricht $\alpha_1 = \ldots = \alpha_k = 3/2$ und der Posteriori-Modus Schätzung

$$\begin{aligned}\text{Modus}(\boldsymbol{\theta}|\boldsymbol{x})_j &= \frac{x_j + \alpha_j - 1}{\sum_{i=1}^k x_i + \sum_{i=1}^k \alpha_i - k} \\ &= \frac{x_j + \frac{1}{2}}{\sum_{i=1}^k x_i + \frac{1}{2}k} \,.\end{aligned}$$

Notation:

- $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ ist p-dimensionaler Zufallsvektor.
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ ist p-dimensionaler Erwartungswertvektor.
- Die symmetrische und positiv definite (Notation: $\boldsymbol{\Sigma} > 0$) Matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{pmatrix}$$

ist $p \times p$-dimensionale Kovarianzmatrix.

- Eine Beobachtung $\boldsymbol{x} = (x_1, \ldots, x_p)^\top$ ist MVN$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (multivariat normalverteilt), wenn

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

- Die Likelihood für n i.i.d. Realisationen $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ lautet

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right) \\ &= |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_0)\right) \end{aligned}$$

mit $\boldsymbol{S}_0 = \sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top$.

...

- ...
  Die zweite Identität ergibt sich über die Umformungen

$$\begin{aligned}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) &= \operatorname{tr}\left(\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right)\\ &= \operatorname{tr}\left(\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top\right)\\ &= \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}_0).\end{aligned}$$

I. Konjugierte Prioriverteilung bei unbekanntem $\boldsymbol{\mu}$ und bekanntem $\boldsymbol{\Sigma}$

Konjugierte **Prioriverteilung** für $\boldsymbol{\mu}$ bei bekanntem $\boldsymbol{\Sigma}$ ist

$$\boldsymbol{\mu} \;\sim\; \mathrm{MVN}(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$$

mit $\boldsymbol{\Lambda}_0 > 0$.
Die **Posteriori** für $\boldsymbol{\mu}$ ist

$$f(\boldsymbol{\mu}|\boldsymbol{x}, \boldsymbol{\Sigma}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Lambda}_0^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right)$$

Der Term im Exponenten ist eine quadratische Form in $\boldsymbol{\mu}$, die sich über eine quadratische Ergänzung und Vernachlässigung von Konstanten ergibt. Man erhält

$$f(\boldsymbol{\mu}|\boldsymbol{x}, \boldsymbol{\Sigma}) \sim \text{MVN}(\boldsymbol{\mu}_n, \boldsymbol{\Lambda}_n)$$

mit

$$\begin{aligned}
\boldsymbol{\mu}_n &= \left(\boldsymbol{\Lambda}_0^{-1} + n\boldsymbol{\Sigma}^{-1}\right)^{-1} \left(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\mu}_0 + n\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{x}}\right), \\
\boldsymbol{\Lambda}_n^{-1} &= \boldsymbol{\Lambda}_0^{-1} + n\boldsymbol{\Sigma}^{-1}
\end{aligned}$$

und $\overline{\boldsymbol{x}} = (\sum_{i=1}^n \boldsymbol{x}_i)/n$ in Analogie zum univariaten Fall.

Die **bedingte** und **marginale Posteriori** für *Subvektoren* von $\boldsymbol{\mu}$ folgen aus den Eigenschaften der multivariaten Normalverteilung: Betrachte die Partitionierungen

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \quad \boldsymbol{\mu}_n = \begin{pmatrix} \boldsymbol{\mu}_n^{(1)} \\ \boldsymbol{\mu}_n^{(2)} \end{pmatrix}, \quad \boldsymbol{\Lambda}_n = \begin{pmatrix} \boldsymbol{\Lambda}_n^{(11)} & \boldsymbol{\Lambda}_n^{(12)} \\ \boldsymbol{\Lambda}_n^{(21)} & \boldsymbol{\Lambda}_n^{(22)} \end{pmatrix}.$$

Dann gilt für die bedingten Verteilungen

$$\boldsymbol{\mu}^{(1)}|\boldsymbol{\mu}^{(2)}, \boldsymbol{x} \;\sim\; \mathsf{MVN}\left(\boldsymbol{\mu}_n^{(1)} + \boldsymbol{\beta}^{1|2}\left(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}_n^{(2)}\right), \boldsymbol{\Lambda}_n^{1|2}\right)$$

mit

$$\begin{aligned}
\boldsymbol{\beta}^{1|2} &= \boldsymbol{\Lambda}_n^{(12)}\left(\boldsymbol{\Lambda}_n^{(22)}\right)^{-1}, \\
\boldsymbol{\Lambda}^{1|2} &= \boldsymbol{\Lambda}_n^{(11)} - \boldsymbol{\Lambda}_n^{(12)}\left(\boldsymbol{\Lambda}_n^{(22)}\right)^{-1}\boldsymbol{\Lambda}_n^{(21)}
\end{aligned}$$

und für die marginalen Verteilungen

$$\boldsymbol{\mu}^{(1)} \;\sim\; \mathsf{MVN}\left(\boldsymbol{\mu}_n^{(1)}, \boldsymbol{\Lambda}_n^{(11)}\right).$$

Die **prädiktive Posterioriverteilung** ist (informell)

$$f(\tilde{\boldsymbol{x}}, \boldsymbol{\mu}|\boldsymbol{x}) = \text{MVN}(\tilde{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot \ \text{MVN}(\boldsymbol{\mu}|\boldsymbol{\mu}_n, \boldsymbol{\Lambda}_n).$$

Im Exponenten erhält man eine quadratische Form in $(\tilde{\boldsymbol{x}}, \boldsymbol{\mu})$. $(\tilde{\boldsymbol{x}}, \boldsymbol{\mu})$ sind gemeinsam multivariat normalverteilt, und daher folgt $\tilde{\boldsymbol{x}}$ als Randverteilung der beiden Komponenten ebenfalls einer multivariaten Normalverteilung. Die erforderlichen Kenngrößen lassen sich über die iterierte Erwartung und Varianz berechnen:

$$\mathbb{E}[\tilde{\boldsymbol{x}}|\boldsymbol{x}] = \mathbb{E}[\mathbb{E}[\tilde{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{x}]|\boldsymbol{x}] = \mathbb{E}[\boldsymbol{\mu}|\boldsymbol{x}] = \boldsymbol{\mu}_n$$

und

$$\begin{aligned}
\mathrm{Var}(\tilde{\boldsymbol{x}}|\boldsymbol{x}) &= \mathbb{E}[\mathrm{Var}(\tilde{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{x})|\boldsymbol{x}] + \mathrm{Var}[\mathbb{E}(\tilde{\boldsymbol{x}}|\boldsymbol{\mu}, \boldsymbol{x})|\boldsymbol{x}] \\
&= \mathbb{E}[\boldsymbol{\Sigma}|\boldsymbol{x}] + \mathrm{Var}[\boldsymbol{\mu}|\boldsymbol{x}] \\
&= \boldsymbol{\Sigma} + \boldsymbol{\Lambda}_n.
\end{aligned}$$

II. Konjugierte Prioriverteilung bei unbekanntem $\boldsymbol{\mu}$ und unbekanntem $\boldsymbol{\Sigma}$

In Abschnitt 4.5.1-II (Folie 409) hatten wir als konjugierte Prioriverteilungen für die Parameter μ und σ^2 der univariaten Normalverteilung

$$\mu|\sigma^2 \ \sim \ N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \quad \text{und} \quad \sigma^2 \ \sim \ \text{inv-}\chi^2(\nu_0, \sigma_0^2),$$

kurz

$$\mu, \sigma^2 \sim \text{N-inv-}\chi^2\left(\mu_0, \frac{\sigma_0^2}{\kappa_0}; \nu_0, \sigma_0^2\right),$$

verwendet.

Hier nun verwenden wir die multivariaten Analoga

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \text{MVN}\left(\boldsymbol{\mu}_0, \frac{1}{\kappa_0}\boldsymbol{\Sigma}\right) \quad \text{und} \quad \boldsymbol{\Sigma} \sim \text{inv-Wishart}_{\nu_0}(\boldsymbol{\Lambda}_0^{-1}),$$

kurz

$$\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \text{MVN-inv-Wishart}\left(\boldsymbol{\mu}_0, \frac{1}{\kappa_0}\boldsymbol{\Lambda}_0; \nu_0, \boldsymbol{\Lambda}_0\right).$$

Die **gemeinsame Prioridichte** ist dann

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\left(\frac{\nu_0+p}{2}+1\right)} \cdot \exp\left(-\frac{1}{2}\left(\text{tr}(\boldsymbol{\Lambda}_0\boldsymbol{\Sigma}^{-1}) - \kappa_0(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right)\right).$$

Dabei bezeichnet ν_0 die a priori Anzahl der Freiheitsgrade, κ_0 die a priori Anzahl an Messungen auf der $\boldsymbol{\Sigma}$-Skala. Die **gemeinsame Posterioriverteilung** von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ lautet

$$\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{x} \sim \text{MVN-inv-Wishart}\left(\boldsymbol{\mu}_n, \frac{1}{\kappa_n} \boldsymbol{\Lambda}_n; \nu_n, \boldsymbol{\Lambda}_n\right).$$

mit

$$\begin{aligned}
\boldsymbol{\mu}_n &= \frac{\kappa_0}{\kappa_0 + n} \boldsymbol{\mu_0} + \frac{n}{\kappa_0 + n} \overline{\boldsymbol{x}}, \\
\kappa_n &= \kappa_0 + n, \\
\nu_n &= \nu_0 + n, \\
\boldsymbol{\Lambda}_n &= \boldsymbol{\Lambda}_0 + \boldsymbol{S} + \frac{\kappa_0 n}{\kappa_0 + n} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^\top,
\end{aligned}$$

wobei $\boldsymbol{S} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top$ in Analogie zum univariaten Fall.

Die Interpretation der Parameter von Folie 416 lässt sich direkt übertragen: Der Posteriori-Erwartungswert ist ein gewichtetes Mittel aus Stichprobenmittelwertvektor und Priori-Erwartungswert. Die Gesamtstreuungsmatrix $\boldsymbol{\Lambda}_n$ lässt sich in Priori-Streuungsmatrix, empirische Streuungsmatrix und Streuung zwischen Priori-Erwartungswert und Stichprobenmittel partitionieren.

Die **marginale Posteriori** für $\boldsymbol{\mu}$ folgt einer multivariaten t-Verteilung mit Parametern $\boldsymbol{\mu}_n$ und $\boldsymbol{\Lambda}_n/(\kappa_n \cdot (\nu_n - p + 1))$, die marginale Posteriori für $\boldsymbol{\Sigma}$ einer inversen Wishart-Verteilung mit Parametern ν_n und $\boldsymbol{\Lambda}_n^{-1}$. Zur Simulation aus der gemeinsamen Posteriori oder aus der prädiktiven Verteilung ist folgender Algorithmus anwendbar:

Algorithmus 7 : Simulation aus der gemeinsamen Posteriori und der prädiktiven Verteilung bei konjugierter Priori

Für $s = 1, \ldots, S$:

1. Ziehe $\boldsymbol{\Sigma}^{(s)}|\boldsymbol{x} \sim \text{inv-Wishart}_{\nu_n}\left(\boldsymbol{\Lambda}_n^{-1}\right)$.
2. Ziehe $\boldsymbol{\mu}^{(s)}|\boldsymbol{\Sigma}^{(s)}, \boldsymbol{x} \sim \text{MVN}\left(\boldsymbol{\mu}_n, \frac{1}{\kappa_n}\boldsymbol{\Sigma}^{(s)}\right)$.
3. Ziehe $\widetilde{\boldsymbol{x}}^{(s)}|\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{x} \sim \text{MVN}\left(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}\right)$.

Dann ist $(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})$ eine Ziehung aus der gemeinsamen Posterioridichte, $\tilde{\boldsymbol{x}}$ eine Ziehung aus der prädiktiven Verteilung.

III. Nichtinformative Prioriverteilung bei unbekanntem $\boldsymbol{\mu}$ und unbekanntem $\boldsymbol{\Sigma}$

Als nichtinformative Prioriverteilung bei unbekanntem $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ eignet sich die multivariate Jeffreys' Priori

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(p+1)/2}.$$

Diese entspricht dem Grenzfall $\kappa_0 \to 0, \nu_0 \to 1, |\boldsymbol{\Lambda}_0| \to 0$ bei der konjugierten Priori. Für die Posteriori-Kenngrößen ergibt sich in diesem Fall für die **bedingte Verteilung** von $\boldsymbol{\mu}$

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{x} \sim \mathrm{MVN}\left(\overline{\boldsymbol{x}}, \frac{1}{n}\boldsymbol{\Sigma}\right),$$

...

...

und für die **marginalen Verteilungen**

$$\begin{aligned} \boldsymbol{\Sigma}|\boldsymbol{x} &\sim \text{inv-Wishart}_{n-1}(\boldsymbol{S}), \\ \boldsymbol{\mu}|\boldsymbol{x} &\sim \text{mv-t}_{n-p}\left(\overline{\boldsymbol{x}}, \frac{1}{n(n-p)}\boldsymbol{S}\right). \end{aligned}$$

Als Beispiel wird im Folgenden die bivariate Normalverteilung betrachtet. Es dient auch dazu, die folgende für die Bayes-Inferenz wichtige Simulationsstrategie, das sogenannte Gibbs-Sampling, zu illustrieren.

Algorithmus 8 : Gibbs-Sampler

Gegeben: Ein mehrdimensionaler stetiger Zufallsvektor $\boldsymbol{X}$ mit Verteilung π. Der Einfachheit halber seien alle Komponenten von $\boldsymbol{X}$ stetig.

Wir erzeugen im Folgenden eine Markovkette $\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)}, \ldots$ mit Startwert $\boldsymbol{X}^{(0)}$ und stationärer Verteilung π. Sei $\boldsymbol{X}^{(t)}$ der aktuelle Zustand der Markovkette. $\boldsymbol{X}^{(t)}$ lasse sich in k Subvektoren $\boldsymbol{X}^{(t)} = (\boldsymbol{X}_{\bullet 1}^{(t)}, \boldsymbol{X}_{\bullet 2}^{(t)}, \ldots, \boldsymbol{X}_{\bullet k}^{(t)})$ partitionieren. Definiere

$$\boldsymbol{X}_{-s}^{(t)} = \left(\boldsymbol{X}_{\bullet 1}^{(t)}, \boldsymbol{X}_{\bullet 2}^{(t)}, \ldots, \boldsymbol{X}_{\bullet (s-1)}^{(t)}, \boldsymbol{X}_{\bullet (s+1)}^{(t-1)}, \ldots, \boldsymbol{X}_{\bullet k}^{(t-1)}\right)$$

für $s = 1, \ldots, k$.

Algorithmus 8 : Gibbs-Sampler fortgeführt

Ferner seien die *vollständig bedingten Verteilungen („full conditionals")*

$$\pi^{(t)}_{\bullet s|-s} = \pi\left(\boldsymbol{x}^{(t)}_{\bullet s} \middle| \boldsymbol{x}^{(t)}_{-s}\right) = \frac{\pi\left(\boldsymbol{x}^{(t)}_{\bullet s}, \boldsymbol{x}^{(t)}_{-s}\right)}{\int \pi\left(\boldsymbol{x}^{(t)}_{\bullet s}, \boldsymbol{x}^{(t)}_{-s}\right) d\boldsymbol{x}^{(t)}_{\bullet s}}$$

gegeben und simulierbar.

Algorithmus 8 : Gibbs-Sampler fortgeführt

Dann wird der nächste Zustand $\boldsymbol{X}^{(t+1)}$ komponentenweise wie folgt erzeugt:

Schritt 1: Ziehe $\boldsymbol{X}_{\bullet 1}^{(t+1)} \sim \pi_{\bullet 1|-1}^{(t+1)}$.

Schritt 2: Ziehe $\boldsymbol{X}_{\bullet 2}^{(t+1)} \sim \pi_{\bullet 2|-2}^{(t+1)}$.

$\vdots$

Schritt k: Ziehe $\boldsymbol{X}_{\bullet k}^{(t+1)} \sim \pi_{\bullet k|-k}^{(t+1)}$.

Wiederhole diese Schritte ausreichend oft.

Nach einer gewissen Zahl von Wiederholungen kann $\boldsymbol{X}^{(t)}$ als Ziehung aus π angesehen werden. Im Gegensatz zu obigen „direkten" Simulationsalgorithmen liegen nun allerdings abhängige Realisationen vor.

Beispiel 4.2. (Bivariate Normalverteilung)
Sei $\boldsymbol{X}$ bivariat normalverteilt mit Erwartungswertvektor $(\mu_1, \mu_2)^\top$ und Kovarianzmatrix

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \text{ und } \rho \text{ bekannt.}$$

Bei einer nichtinformativen Priori $p(\mu_1, \mu_2) \propto \text{const}$ für μ_1, μ_2 reduziert sich die Posteriori auf die Likelihood bei gegebenen Daten $\boldsymbol{x} = ((x_{11}, x_{12})^\top, \ldots, (x_{n1}, x_{n2})^\top)$:

$$L(\mu_1, \mu_2) = \left(\frac{1}{2\pi}\right)^n \left(1 - \rho^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2(1-\rho^2)} A\right),$$

wobei

$$A = \sum_{i=1}^{n} \left[(x_{i1} - \mu_1)^2 - 2\rho(x_{i1} - \mu_1)(x_{i2} - \mu_2) + (x_{i2} - \mu_2)^2\right]$$

...

...

Wir möchten nun die vollständig bedingten Verteilungen $\mu_1|\mu_2, \boldsymbol{x}$ und $\mu_2|\mu_1, \boldsymbol{x}$ berechnen. Natürlich ist es aus Symmetriegründen ausreichend, nur $\mu_1|\mu_2, \boldsymbol{x}$ zu ermitteln. Wegen

$$f(\mu_1|\mu_2, \boldsymbol{x}) = \frac{f(\mu_1, \mu_2|\boldsymbol{x})}{f(\mu_2|\boldsymbol{x})} = \frac{f(\mu_1, \mu_2, \boldsymbol{x})}{f(\mu_2, \boldsymbol{x})} \propto f(\mu_1, \mu_2|\boldsymbol{x})$$

genügt es, aus der gemeinsamen Posteriori lediglich die Terme zu betrachten, die von der jeweiligen Variablen in der bedingten Verteilung abhängen. Man erhält dann

$$f(\mu_1|\mu_2, \boldsymbol{x}) \propto \exp\left(-\frac{1}{2(1-\rho^2)}\, n\left[\mu_1^2 - 2\mu_1(\overline{x}_1 + \rho(\mu_2 - \overline{x}_2))\right]\right)$$

mit $\bar{x}_j = (\sum_{i=1}^n x_{ij})/n$ für $j = 1, 2$.

...

Eine quadratische Ergänzung des Terms in eckigen Klammern um $\overline{x}_1 + \rho(\mu_2 - \overline{x}_2)$ liefert schließlich das Endresultat

$$p(\mu_1|\mu_2, \boldsymbol{x}) \propto \exp\left(-\frac{1}{2\frac{1-\rho^2}{n}}\left(\mu_1 - [\overline{x}_1 + \rho(\mu_2 - \overline{x}_2)]\right)^2\right).$$

Dies entspricht dem Kern einer $N(\overline{x}_1 + \rho(\mu_2 - \overline{x}_2), (1-\rho^2)/n)$-Verteilung. Der zugehörige Gibbs-Sampler hat die Gestalt:

1. Wähle einen Startwert $\mu_2^{(0)}$.
2. Für $s = 1, \ldots, S$:
   (a) Ziehe $\mu_1^{(s)}|\mu_2^{(s-1)} \sim N\left(\overline{x}_1 + \rho\left(\mu_2^{(s-1)} - \overline{x}_2\right), \frac{1-\rho^2}{n}\right)$.
   (b) Ziehe $\mu_2^{(s)}|\mu_1^{(s)} \sim N\left(\overline{x}_2 + \rho\left(\mu_1^{(s)} - \overline{x}_1\right), \frac{1-\rho^2}{n}\right)$.