Bayesova statistika - zapiski s predavanj prof. Smrekarja

Tomaž Poljanšek

študijsko leto 2023/24

Kazalo

Seznam uporabljenih kratic

kratica	pomen
s.v.	slučajni vektor
B	binomska porazdelitev
NEP	neodvisen in enako porazdeljen
s.s.	slučajna spremenljivka
p.v.	pričakovana vrednost
AKI	aposteriorni kredibilnostni interval
BF	Bayesova formula
s.g.	skoraj gotovo
k.p.f.	kumulativna porazdelitvena funkcija
A/R	accept or reject
M.v.	Markovska veriga
EZVŠ	ergodični zakon veliikih števil
ECLI	ergodični CLI
M-H	Matropolis-Hasting
HM	hierarhični model
EIZ	ergodični interval zaupanja
MCMC	Monte Carlo markovske verige
SNK	?
MNK	metoda najmanjših kvadratov
VKR	vsota kvadratov residualov
UNK	uteženi najmanjši kvadrati

Poglavje 1

Uvod

Bayesova statistika je formalni okvir za „osveževanje“ vedenja/znanja o porazdelitvi nekega slučajnega vektorja.

Zgled. 1000, $\approx$ 400Č $\rightarrow$ 600B (apriorno znanje).
Izvedemo (statistični) poskus: izvlečemo 10, dobimo 6 črnih in 4 bele

1.1 Elementarna Bayesova statistika

Privzamemo popoln sistem dogodkov $E_1, E_2 \ldots E_m : E_i \cap E_j = \emptyset$ za $i \neq j$ in $E_1 \cup E_2 \cup \cdots \cup E_m = \Omega$.
Če imamo še neki dogodek A, velja t.i. zakon o popolni verjetnosti
$P(A) = \sum_{i=1}^{m} P(A \mid E_i) \cdot P(E_i)$ (interpretacija: 2-fazni poskus).
V Bayesovem okviru nas zanimajo $P(E_j \mid A)$ (verjetnost, da se je v „1. fazi“ zgodil E_j, če se je „2. fazi“ zgodil A). Ker je

$$P(E_j \mid A) = \frac{P(E_j \cap A)}{P(A)}$$

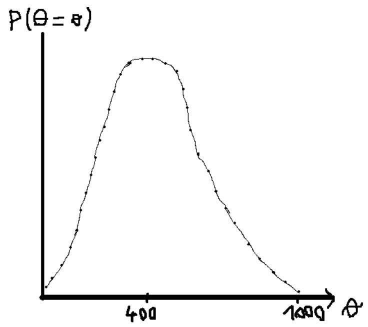
je

$$P(E_j \mid A) = \frac{P(A \mid E_j) \cdot P(E_j)}{P(A)} \quad \text{- elementarna pogojna verjetnost}$$

oziroma

$$P(E_j \mid A) = \frac{P(A \mid E_j) \cdot P(E_j)}{\sum_{i=1}^{m} P(A \mid E_i) \cdot P(E_i)} \quad \text{- elementarna Bayesova formula.}$$

Nadaljujemo zgled. V Bayesovi statistiki predhodno („apriorno“) vedenje formaliziramo kot realizacijo slučajnega eksperimenta. V našem primeru vpeljemo fukcijo, da smo število črnih frnikul θ (- realizacija) dobili kot rezultat slučajne spremenljivke $\Theta \in \{0, 1, 2 \ldots 1000\}$.
Informacijo $\theta \approx 400$ zakodiramo kot $E(\Theta) = 400$.



Privzamemo (kar!) $\Theta \sim B\left(1000, \frac{4}{10}\right)$
$\Longrightarrow \;\; P(\Theta = \theta) = \binom{1000}{\theta} \left(\frac{4}{10}\right)^{\theta} \left(1 - \frac{4}{10}\right)^{1000-\theta}.$
$P(k$ črnih od 10 izvlečenih$|\ \Theta = \theta) = \frac{\binom{\theta}{k}\binom{1000-\theta}{10-k}}{\binom{10}{k}}$ (*)
(*) pri omejitvah (k omejimo).
Osvežena porazdelitev - novo vedenje

$$P(\Theta = \theta \mid 6 \text{ črnih od 10 izvlečenih}) = \\ \frac{P(6 \text{ črnih od 10 izvlečenih} \mid \Theta = \theta) \cdot P(B(1000, \frac{4}{10}) = \theta)}{\sum_{i=0}^{1000} P(6 \text{ črnih od 10 izvlečenih} \mid \Theta = i) \cdot P(B(1000, \frac{4}{10})) = i}.$$

Pravimo ji aposteriorna porazdelitev.

1.2 Proučevani slučajni vektor (vzročni) parametrični model

Naj bo $X = (X_1, X_2 \dots X_n) \in \mathbb{R}^n$ preučevani slučajni vektor. Pogosto so neodvisni in enako porazdeljeni (NEP) realizacija danega slučajnega eksperimenta. S pomočjo statistike lahko „ocenjujemo“ porazdelitev slučajnega vektorja X. Zanjo privzamemo, da pripada nekemu modelu, t.j. neki množici dopustnih rešitev. Privzamemo, da je ta množica parametrizirana s parametričnim prostorom $\Theta \subset \mathbb{R}^r$. Tu si mislimo, da parameter $\theta \in \Theta$ dobimo kot realizacijo slučajnega vektorja (s.v.) Θ z vrednostmi v Θ (večinoma $r \geq 2$). Porazelitvi s.v. X_i pogojno na $\Theta = \theta$ pravimo vzorčna porazdelitev. Privzeli bomo, da imamo gostote $f(x \mid \theta)$ ali verjetnostne funkcije

$$P(X = x \mid \theta) = f(x \mid \theta),$$

torej da velja

$$P(X \in B \mid \Theta = \theta) = \int_B f(x \mid \theta) d\nu(x)$$

(v Lebesgueovi meri) ali

$$P(X \in B \mid \Theta = \theta) = \sum_{x \in B} f(x \mid \theta).$$

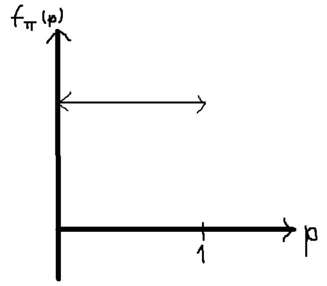
Modelu pogojnih porazdelitev $(X \mid \Theta = \theta)$ pravino vzorčni model.

1.3 Apriorna in „robna“ porazdelitev

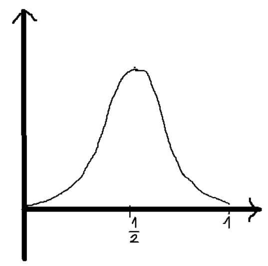
Porazdelitvi fiktivnega slučajnega vektorja Θ pravimo apriorna porazdelitev, brezpogojni (robni) porazdelitvni slučajnega vektorja X pa pravimo „robna“ porazdelitev
(*) v resnici sta obe porazdelitvi robni porazdelitvi družne porazdelitve vektorja (X, Θ) z vrednostmi v $\mathbb{R}^{n+r}$.

Zgled. Ocenjujemo Bernoullijevo porazdelitev. Predhodno vedenje je podano z apriorno prazdelitvijo na (0,1); mislimo si, da je p realizacija slučajne spremenljivke (s.s.) Π z vrednostmi v (0,1). Možnosti:

- nimamo apriornega mnenja o (dejanskem) p: tedaj bi (morda) vzeli zvezno porazdelitev z gostoto enakomerna porazdelitve,



- smo „zelo“ prepričani, da je (dejanski) $p \approx \frac{1}{2}$.



Recimo, da je $f(p)$ gostota apriorne porazdelitve. Tedaj so apriorne verjetnosti

$$P(\Pi \in (a,b)) = \int_a^b f(p)dp$$

in apriorna pričakovana vrednost

$$E(\Pi) = \int_0^1 pf(p)dp.$$

Pripomnimo, da pri $\Pi \sim U(0,1)$ dobimo $E(U(0,1)) = \frac{1}{2}$.
Privzemimo, da smo „vzorčili“ p, potem pa „neodvisno“ n-krat vržemo p-kovanec (P(cifra$= p$)), gre za slučajne spremenljivke $X_1, X_2 \ldots X_n$, za katere

je $(X_i \mid \Pi = p) \sim Bernoulli(p)$ in so $X_1 \dots X_n$ neodvisne pogojno na p. To ne pomeni, da do $X_1 \dots X_n$ brezpogojno neodvisne.
Za $i \neq j$ je

$$
\begin{aligned}
P(X_i = 1 \wedge X_j = 1) = \int_0^1 & P(X_i = 1 \wedge X_j = 1 \mid \Pi = p) f(p) dp = \\
& \overset{\text{pogojno neodvisne}}{=} \int_0^1 P(X_i = 1 \mid \Pi = p) P(X_j = 1 \mid \Pi = p) f(p) dp = \\
& = \int_0^1 p^2 f(p) dp = \\
& = E(\Pi^2).
\end{aligned}
$$

Ker je $P(X_i = 1) = \int_0^1 P(X_i = 1 \mid p) f(p) dp = \int_0^1 p f(p) dp = E(\Pi)$, je

$$
Cov(X_i, X_j) = E(\Pi^2) - E(\Pi)^2 = D(\Pi)
$$

za $i \neq j$, torej so X_i brezpogojno neodvisne $\iff$ $\Pi =$ konstantna (slučajna spremenljivka).
Tvorimo $X = X_1 + \dots + X_n \in \{0, 1 \dots n\}$. To je „preučevana“ slučajna spremenljivka. Velja $(X \mid \Pi = p) \sim B(n,p)$. To je vzorčna porazdelitev; vzročni model je parametriziran s prostorom parametrov $(0,1) = \Theta$. Robna porazdelitev je podana z verjetnostmi

$$
\begin{aligned}
P(X = k) &= \int_0^1 P(X = k \mid p) f(p) dp = \\
&= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p) dp.
\end{aligned}
$$

Recimo, da „opazimo“ $X = k$. Aposteriorna porazdelitev (osveženo vedenje o p) je sestavljeno iz verjetnosti

$$
\begin{aligned}
P(X \in (a,b) \mid X = k) &= \frac{P(X = k \wedge \Pi \in (a,b))}{P(X = k)} = \\
&= \frac{\int_0^1 P(X = k \wedge \Pi \in (a,b) \mid \Pi = p) f(p) dp}{P(X = k)} = \\
&= \int_a^b \frac{P(X = k \mid \Pi = p)}{P(X = k)} f(p) dp.
\end{aligned}
$$

Opazimo, da ima aposteriorna porazdelitev $(\Pi \mid X = k)$ gostoto

$$f_{(\pi|X)}(p \mid k) = \frac{P(X = k \mid p)f(p)}{P(X = k)}.$$

Zgornji formuli pravimo Bayesova formula.
Za številsko oceno za p bi lahko vzeli pričakovano vrednost aposteriorne porazdelitve

$$\hat{p} = E(\Pi \mid X = k) = \int_0^1 p \cdot f(p \mid k)dp.$$

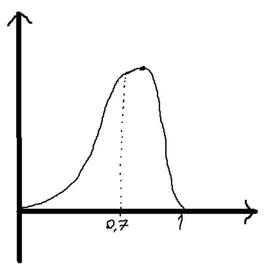
Pravimo ji aposteriorna pričakovana vrednost.
Posebej priročna družina apriornih porazdelitev (za binomske vzorčne porazdelitve) je t.i. $Beta = \{Beta(a,b) \mid a,b \in (0,\infty)\}$

$$f_{Beta(a,b)}(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} 1_{(0,1)}(p)$$

(tu je $B(a,b) = \int_0^1 p^{a-1}(1-p)^{b-1}dp$).

$$E(Beta(a,b)) = \frac{a}{a+b}$$
$$D(Beta(a,b)) = \frac{ab}{(a+b)^2(a+b+1)}.$$

$D(Beta(a,b))$ predstavlja „težo“ apriornega prepričanja; večji - manj sigurni smo.



$$E(Beta(a,b)) = 0.7.$$

Aposteriorna porazdelitev ima gostoto (če je $f(p) = f_{Beta(a,b)}(p)$)

$$f(p \mid k) = \frac{\binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}}{P(X=k)} = \\ = \text{konst.} \cdot p^{a+k-1}(1-p)^{b+n-k-1}.$$

Vidimo, da je $(\Pi \mid X = k) \sim Beta(a+k, b+n-k)$.
Aposteriorna pričakovana vrednost (p.v.) je

$$\frac{a+k}{a+b+n} = \frac{(a+b)\frac{a}{a+b} + n\frac{k}{n}}{a+b+n} = \\ = \frac{a+b}{a+b+n} \cdot \frac{a}{a+b} + \frac{n}{a+b+n} \cdot \frac{k}{n}.$$

Tukaj je

- $\frac{a}{a+b}$ apriorna ocena,
- $\frac{k}{n}$ vzorčna ocena in
- $\frac{a+b}{a+b+n}$ in $\frac{n}{a+b+b}$ faktorja pri konveksni kombinaciji obeh ocen.

Vzorec velik $\to$ prevlada mnenje vzorca.

1.4 Disperzija aposteriornih porazdelitev

Gre pravzaprav za disperzijo pogojnih porazdelitev. Naj bosta $X : \Omega \to \mathbb{R}^m$ in $Y : \Omega \to \mathbb{R}^n$ in naj ima (X,Y) gostoto $f_{(X,Y)}$ glede na $\mu \times \nu$ Sledita gostoti $f_X(x) = \int f_{(X,Y)}(x,y) d\nu(y)$ za X glede na μ in
$f_Y(y) = \int f_{(X,Y)}(x,y) d\mu(x)$ za Y glede na ν. Dalje definiramo pogojni porazdelitvi $(Y \mid X = x)$ in $(X \mid Y = y)$ preko gostot

$$f_{(Y|X)}(y \mid x) = \frac{f_{(X,Y)}(X,Y)}{f_X(x)}$$

glede na ν: gostota v $X \to \mu$ in simetrično za $f_{(X|Y)}(x \mid y)$.
$P(Y \in B) \mid X = x = \int_B f_{(Y|X)}(y \mid x) d\nu(y)$ - porazdelitev, opremljena z gostoto.

Definicija 1.4.1.

$$E(Y \mid X = x) = \int y f_{(Y|X)}(y \mid x) d\nu(y).$$

y lahko zamenjamo s $h(y)$.
Pišemo $E(Y \mid X = x) = u(X)$ - h je identiteta.

Definicija 1.4.2.

$$E(Y \mid X) = u(X) : \Omega \to \mathbb{R}^n.$$

Slučajni vektor $\to$ pogojna pričakovana vrednost,
oz.

$$E(Y \mid X)(\omega) = u(X(\omega)) = E(Y \mid X = X(\omega)).$$

$E(Y \mid X)(\omega)$: funkcija na X, kompozitum.
$X(\omega)$: vrednost.

Definicija 1.4.3. Pogojno varianco slučajnega vektorja Y, pogojno na $X = x$ definiramo kot varianco pogojne porazdelitve $(Y \mid X = x)$, t.j.

$$E((Y - u(X))(Y - u(X))^T \mid X = x) =: Var(Y \mid X = x).$$

Ker je E aditivna, velja

$$E((Y - u(X))(Y - u(X))^T \mid X = x) = E(YY^T \mid X = x) - u(X)u(X)^T =: v(X).$$

$v(X)$ je $n \times n$ matrika.

Definicija 1.4.4. Pogojna varianca slučajnega vektorja Y pogojno na slučajni vektor X je

$$Var(Y \mid X) = v(X).$$

Zadnjič: beta-binomski model: proučevana s.s. T; vzorčne porazdelitve $(T \mid p) \sim B(n,p)$ za $p \in \Theta = (0,1)$. Če je apriorna porazdelitev $Beta(a,b)$, je aposteriorna pri $T = k$ enaka $Beta(a + k, b + n - k)$.
Disperzija porazdelitve $(\Pi \mid T = k)$ je enaka $\frac{(a-k)(b+n-k)}{(a+b+k)^2(a+b+n+1)}$ (in je odvisna od realizacije k).

DN: za katere k je $D(\text{aposteriorne}) > D(\text{apriorne})$.

Izkaže se, da je za nekatere k manjša, za nekatere k pa večja od disperzije apriorne porazdelitve. Vendar pa velja t.i. „zakon popolne porazdelitve“

$$Var(\Theta) = E(Var(\Theta \mid X)) + Var(E(\Theta \mid X)),$$

kjer sta seveda

$$E(Var(\Theta \mid X)) \geq 0 \text{ in}$$
$$Var(E(\Theta \mid X)) \geq 0,$$

(X vzorčni), iz katerega sledi $E(Var(\Pi \mid X)) \leq Var(\Theta)$ (apriori). V konkretnem primeru je

$$E(Var(\Pi \mid T)) \leq Var(\Pi).$$

Opomba.

$$E(Var(\Pi \mid T)) = \sum_{k=0}^{n} Var(\Pi \mid k) \cdot P(T = k);$$

T vzorčimo, T je robna porazdelitev (binomska pogojna?).

(V povprečju aposteriorna varianca boljša.)

1.5 Aposteriorni kredibilnostni interval

(Bayesova različica intervala zaupanja).

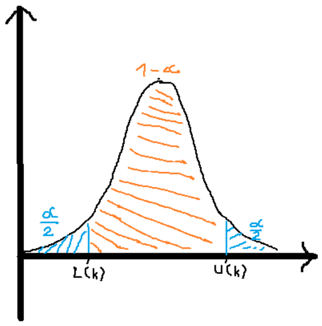
V konkretnem zgledu „iščemo“ funkciji realizacije $L(k), U(k)$, za kateri velja

$$\forall k \; P\left(L(k) \leq p \leq U(k)\right) \geq 1 - \alpha.$$

p je slučajen.

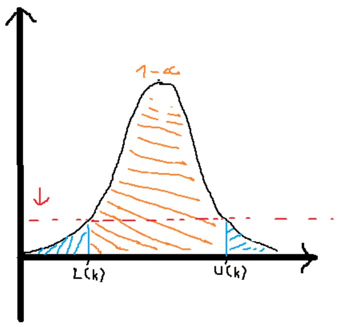
Ker za $\Pi \sim Beta(a,b)$, vemo $(\Pi \mid T = k) \sim Beta(a + k, b + n - k)$, lahko izberemo

$$L(k) = F^{-1}_{Beta(a+k,b+n-k)}\left(\frac{\alpha}{2}\right),$$
$$U(k) = F^{-1}_{Beta(a+k,b+n-k)}\left(1 - \frac{\alpha}{2}\right)$$



in v aposteriorni kredibilnostni interval (AKI) dobimo $= 1-\alpha$. Taki konstrukciji pravimo centralni kredibilnostni interval. V praksi pogosto uporabljamo tudi t.i. kredibilnostni interval največje gostote, kjer zahtevamo

$$f_{(\Pi|T=k)}(L(k)) = f_{(\Pi|T=k)}(U(k)).$$



To ima smisel za unimodalne aposteriorne porazdelitve.

1.6 Splošne oznake

X - proučevalni vektor (z vrednostmi v $\mathbb{R}^n$).

x - realizacija vektorja $X(x \in \mathbb{R}^n)$.

$\Theta \subset \mathbb{R}^n$ - parametrični prostor.
Θ - (fiktivni) vektor z realizacijo $\theta \in \Theta$.
$P = \{P_\theta = (X \mid \Pi = \theta) \mid \theta \in \Theta\}$ - družina vzorčnih porazdelitev (vzorčni model).
$f(x \mid \theta)$ - gostota (ali verjetnostna funkcija) porazdelitve $(X \mid \Theta = \theta)$ (izračunano v x).

Opomba. $f(x \mid \theta) = f_{X|\Theta}(x \mid \theta)$ - spuščamo.

V istem smislu gostota (ali verjetnostna funkcija) apriorne porazdelitve. Za aposteriorno gostoto (ali verjetnostno funkcijo) velja Bayesova formula (BF)

$$f(\theta \mid x) = \frac{f(x \mid \theta) f(\theta)}{f(\theta)} \propto f(x \mid \theta) f(\theta).$$

$f(x)$ je normalizacijski faktor.

Poglavje 2

Enoparametrični modeli

2.1 Beta-binomski model

Uvodni beta-binomski model je enoparametričen.

2.2 Poissonov model (gama-poissonov model)

Naj bo parametrični prostor $\Theta = (0, \infty)$ in naj bo $X = (X_1 \dots X_n)$, kjer so $(X_i \mid \lambda) \overset{\text{NEP}}{\sim} Poisson(\lambda)$. Za proučevano s.s. vzamemo $T = \sum_{i=1}^{n} X_i$; seveda je $(T \mid \lambda) \sim Poisson(n\lambda)$.
Privzemimo apriorno porazdelitev

$$f(\lambda) = f_{Gama(a,b)}(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \cdot 1_{(0,\infty)}(\lambda).$$

(0,∞) je parametrični prostor, a in b sta pozitivni konstanti.
Izkaže se:

$$E(Gama(a,b)) = \frac{a}{b},$$
$$D(Gama(a,b)) = \frac{a}{b^2}.$$

DN.

$$P(T = k \mid \lambda) = e^{-n\lambda} \frac{(n\lambda)^k}{k!}.$$

Bayesova formula se glasi

$$\begin{aligned} f(\lambda \mid T = k) &\propto P(T = k \mid \lambda) \cdot f(\lambda) \\ &\propto e^{-n\lambda} \lambda^k \cdot \lambda^{a-1} e^{-b\lambda} \\ &= \lambda^{a+k-1} e^{-(b+n)\lambda} \end{aligned}$$

$\implies$ $(\Lambda \mid T = k) \sim \mathit{Gama}(a + k, b + n)$.

Definicija 2.2.1. Naj bo podan vzorčni model P in naj bo K družina porazdelitev na parametričnem prostoru Θ. Pravimo, da je K konjugirana k P, če vedno velja

$$f(\theta \mid x) \in K \implies \forall x : \ (\Theta \mid X = x) \in K.$$

$f(\theta \mid x)$ je porazdelitev na Θ. Rečemo lahko tudi, da sta P in K konjugiran par.

2.3 Normalni model z znano disperzijo

Tu je σ^2 znana disperzija, vzorec $X = (X_1 \dots X_n)$ pa zadošča $(X_i \mid \mu) \overset{\text{NEP}}{\sim} N(\mu, \sigma^2)$, kjer je $\mu \in \Theta = \mathbb{R}$. S katero porazdelitvijo bi zakodirali apriorno informacijo?

Recimo, da je apriorno mnenje: $\mu \approx \mu_0$. Vzemimo za apriorno porazdelitev kar $N(\mu_0, \tau_0^2)$.

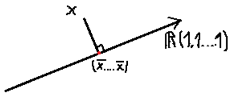
Vzorčna:

$$\begin{aligned} f(x \mid \mu) &= f(x_1 \dots x_n \mid \mu) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}. \end{aligned}$$

Pripomnimo, da je

$$\sum_{i=1}^n (x_i - \mu)^2 = n \cdot (\mu - \overline{x})^2 + \sum_{i=1}^n (x_i - \overline{x})^2,$$

kjer je $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.



$\mathbb{R}(1,1\ldots1)$ - prostorska diagonala.
(Vzorčna: jih je več, apriorna: ena.)
Apriorna:

$$f(\mu) = (2\pi\sigma_0^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}.$$

Opazimo, da je

$$f(\mu) = e^{\text{kvadratni polinom}(\mu)}.$$

Velja:

$$\int_{-\infty}^{\infty} e^{a\mu^2+b\mu+c} d\mu < \infty \iff a < 0.$$

DN (kako se narobe lotiti): drugače koliko apriorna gostota (integral robne gostote, Bayesova formula).
V tem duhu

$$f(\mu \mid x) \propto e^{-\frac{1}{2}\left(\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)\mu^2 - 2\left(\frac{n\overline{x}}{\sigma^2}-\frac{\mu_0}{\tau_0^2}\right)\mu\right)}.$$

Prepoznamo kot normalno porazdelitev. Označimo jo $N(\mu_1, \tau_1^2)$.
Velja

$$f_{N(\mu_1,\tau_1^2)}(\mu) \propto e^{-\frac{1}{2}\left(\frac{1}{\tau_1^2}\mu^2 - 2\frac{\mu_2}{\tau_1^2}\mu\right)}.$$

Sledi

$$[\mu^2] : \ \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} = \frac{1}{\tau_1^2} \text{ in}$$
$$[\mu] : \ \frac{n\overline{x}}{\sigma^2} + \frac{\mu_0}{\tau_0^2} = \frac{\mu_1}{\tau_1^2}.$$

Tukaj je:

- $\frac{n}{\sigma^2}$ vzorčna preciznost,
- $\frac{1}{\tau_0^2}$ apriorna preciznost,

- $\frac{1}{\tau_1^2}$ aposteriorna preciznost,
- preciznosti se pri seštevanju neodvisnih normalnih porazdelitvah seštevajo.

Preciznost je $\frac{1}{D(..)}$.
To pomeni

$$\tau_1^2 = (\frac{n}{\sigma^2} + \frac{1}{\tau_0^2})^{-1}$$

in

$$\begin{aligned}
\mu_1^2 &= \tau_1^2(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}) \\
&= \frac{(\frac{n}{\sigma^2}\overline{x} + \frac{1}{\tau_0^2}\mu_0) \cdot \frac{\tau_0^2\sigma^2}{n}}{(\frac{n}{\sigma^2} + \frac{1}{\tau_1^2}) \cdot \frac{\tau_0^2\sigma^2}{n}} \\
&= \frac{\tau_0^2\overline{x} + \frac{\sigma^2\mu_0}{n}}{\tau_0^2 + \frac{\sigma^2}{n}} \\
&= \frac{\tau_0^2}{\tau_0^2 + \frac{\sigma^2}{n}}\overline{x} + \frac{\frac{\sigma^2}{n}}{\tau_0^2 + \frac{\sigma^2}{n}}\mu_0
\end{aligned}$$

2.4 Eksponentne družine porazdelitev

Vzorčni model pripada eksponentni družini porazdelitev, če velja

$$\begin{aligned}
f(x \mid \theta) &= c(\theta) \cdot e^{\langle Q(\theta), T(x) \rangle} \cdot h(x) \\
&= e^{-\psi(\theta)} e^{\langle Q(\theta), T(x) \rangle} \cdot h(x), \qquad (2.1)
\end{aligned}$$

kjer je

$$\begin{aligned}
&\tau : \Theta \to \mathbb{R} \\
&T : \mathbb{R}^n \to \mathbb{R}^m \\
&Q : \Theta \to \mathbb{R}^m \text{ in} \\
&h : \mathbb{R}^n \to [0, \infty].
\end{aligned}$$

Zgled.

(1) NEP Bernoullijev model:

$$(X_i \mid p) \overset{\text{NEP}}{\sim} Bernoulli(p) = B(1{,}p)$$
$$f(x_1 \ldots x_n \mid p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \cdot 1_{\{0,1\}^n}(x_1 \ldots x_n)$$

$$f(x_1 \ldots x_n \mid p) = P(X_1 = x_1 \ldots X_n = x_n \mid p)$$

Preoblikujemo v:

$$(1-p)^n \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} \cdot 1_{\{0,1\}^n}(x_1 \ldots x_n) = e^{\ln\left(\frac{p}{1-p}\right)\sum_{i=1}^n x_i} \cdot 1_{\{0,1\}^n}(x_1 \ldots x_n).$$

(2) Normalni model z znano σ^2:
$\Theta = \mathbb{R}, \mu \in \mathbb{R}$

$$\begin{aligned}(EXP)\ f(x_1 \ldots x_n \mid \mu) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2}.\end{aligned}$$

Tukaj je

$$\begin{aligned} c(\mu) &= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{n\mu^2}{2\sigma^2}} \\ Q(\mu) &= e^{-\frac{\mu}{\sigma^2}} \\ T(x) &= e^{\sum_{i=1}^n x_i} \\ h(x) &= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2} \end{aligned}$$

(3) Normalni model z neznano disperzijo:
$\Theta = \mathbb{R} \times (0,\infty) = (\mu, \sigma^2)$
Zapišemo

$$f(x_1 \ldots x_n \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}} e^{\left\langle \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right), \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)\right\rangle}.$$

Tukaj je

$$
\begin{aligned}
c(\mu) &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\mu^2}{2\sigma^2}} \\
Q(\mu) &= \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right) \\
T(x) &= \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right).
\end{aligned}
$$

$m = 2$.
Preimenujemo (EXP) in definirajmo apriorne gostote

$$f(\eta, \upsilon) = \frac{1}{K(\eta, \upsilon)} \cdot e^{\langle Q(\theta), \eta\rangle - \upsilon\psi(\theta)}. \tag{2.2}$$

Tukaj je $\mu \in \mathbb{R}^n$ in $\upsilon \in \mathbb{R}$.
(Upoštevati moramo morebitne omejitve zaradi zahteve $\int F = 1$).
Seveda je

$$K(\eta, \upsilon) = \int e^{\langle Q(\theta), \upsilon\rangle - \upsilon\psi(\theta)} d\theta.$$

Aposteriorna gostota?

$$f(\theta \mid x) \propto e^{-(\upsilon+1)\psi(\theta) + \langle Q(\theta), \eta + T(x)\rangle}$$

Tukaj smo zmožili nekonstantne faktorje iz 2.1 in 2.2.
Vidimo:

$$f(\theta \mid x) = f_{(\eta + T(x), \upsilon + 1)}(\theta).$$

Gre za konjugirano družino.

Zgled. Aplicirajmo to konstrukcijo na modelu 2. Dobimo konjugirano družino

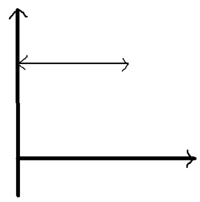
$$f_{(\eta,\upsilon)}(\mu) \propto e^{\frac{\mu}{\sigma^2 \eta} - \tau \frac{n\mu^2}{2\sigma^2}},$$

kjer sta $\eta, \upsilon \in \mathbb{R}$.
Vidimo, da mora biti $\upsilon > 0$.
DN: $\eta, \upsilon \to \mu_0, \tau_0^2$ - reparametrizacija.

2.5 Neinformirane apriorne porazdelitve

Neinformirana apriorna porazdelitev je taka, ki „ne vsebuje predhodne informacije“ o parametru. Tovrsten koncept je šibko-informativna apriorna porazdelitev. Izkaže se, da s statistični praksi (tudi v aplikacija frekventistične statistike) potrebujemo ta koncept.

Zgled. V Beta-binomskem modelu (..) je Laplace predlagal $U(0,1) = Beta(1,1)$ kot neinformirano porazdelitev.



„Ploščata porazdelitev“.

Učinek reparametrizacije

Zgled. Binomski vzorčni model: $f(k \mid p) = \binom{n}{k} p^k (1-p)^{n-k},\ p \in (0,1)$. Reparametrizirajmo s parametrom $q = \ln\left(\frac{p}{1-p}\right) = logit(p)$. Dobimo

$$\begin{aligned}
\widetilde{f}(k \mid q) &= f(k \mid logit^{-1}(q)) \\
&= f\left(k \mid \frac{e^q}{1+e^q}\right) \\
&= \binom{n}{k} \left(\frac{e^q}{1+e^q}\right)^k \left(\frac{1}{1+e^q}\right)^{n-k} \\
&= \binom{n}{k} (1+e^q)^{-k} e^{kq} \quad (q \in \mathbb{R}).
\end{aligned}$$

Kako je s transformacijo apriorne porazdelitve?

Naj bo Π slučajna spremenljivka z realizacijo p in Q slučajna spremenljivka z realizacijo q; velja $Q = logit \circ \Pi = logit(\Pi)$:

$$f_Q(q) = f_{logit(\Pi)}(q) = f_\Pi(logit^{-1}(q)) \cdot \left| \frac{d}{dq} logit^{-1}(q) \right|$$
$$logit^{-1}(q) = 1 - \frac{1}{1+e^q} \implies \frac{d}{dq} logit^{-1}(q) = (1+e^q)^{-2} e^q.$$

Sledi: če je $\Pi \in (0{,}1)$, je

$$f_Q(q) = \frac{e^q}{(1+e^q)^2};$$

ali je to še ploščata porazdelitev?

Jeffreys je kot privzeto neinformativno porazdelitev predlagal

$$f(\theta) \propto \sqrt{|det(FI(\theta))|}. \tag{2.3}$$

Tu je $FI(\theta)$ t.i. Fisherjeva informacija:

$$\begin{aligned} FI(\theta) &= E_{(X|\Theta)}(grad_\theta \ln(f(x \mid \theta)) \cdot grad_\theta \ln(f(x \mid \theta))^T) \\ &= Var_{(X|\Theta)}(grad_\theta \ln(f(x \mid \theta))) \\ &= -E_{(X|\Theta)}(H_\theta(lnf)(x \mid \theta)). \end{aligned}$$

Za $\theta \in \Theta \subset \mathbb{R}^r$ je $grad_\theta \ln(f(x \mid \theta))$ vektor s komponentami $\frac{\partial}{\partial \theta_i} \ln(f(x \mid \theta))$ za $1 \leq i \leq r$.

Učinek reparametrizacije na Jeffreysovo apriorno porazdelitev

Reparametrizirajmo s parametrom $\lambda = \phi(\theta)$, kjer je

$$\phi : \Theta \subset \mathbb{R}^n \to \Lambda \subset \mathbb{R}^n$$

diferenciabilen; sledi

$$\widetilde{f}(x \mid \lambda) = f(x \mid \phi^{-1}(\lambda)).$$

Odvajajmo

$$\begin{aligned} \frac{\partial}{\partial \lambda_i} \ln\Big(\widetilde{f}(x \mid \lambda)\Big) &= \sum_{j=1}^{n} (\frac{\partial}{\partial \lambda_j} \ln(f(x \mid \lambda_j)) \cdot \frac{\partial (\psi^{-1})_j}{\partial \lambda_i}(\lambda)) \\ &= \left[\frac{\partial (\psi^{-1})_j}{\partial \lambda_i} \right]_{j=1}^{n} \cdot grad_\theta(\ln f)(x \mid \phi^{-1}(\lambda)) \end{aligned}$$

$$grad_\lambda f(x \mid \lambda) = [J(\phi^{-1}(\lambda))]^T \cdot grad_\theta \ln f(x \mid \phi^{-1}(\lambda))$$

J: Jacobijeva matrika.

$$\begin{aligned}
\ln \widetilde{f}(x \mid \lambda) &= \ln\big(f(x \mid \phi^{-1}(\lambda))\big) \cdot grad_\lambda \ln(f(x \mid \lambda)) \cdot grad_\lambda \ln(f(x \mid \lambda))^T \\
&= [J\phi^{-1}(\lambda)]^T \cdot grad_\theta \ln f(x \mid \phi^{-1}(\lambda)) \cdot grad_\theta \ln f(x \mid \phi^{-1}(\lambda))^T \cdot [J\phi^{-1}(\lambda)] \\
&\implies \widetilde{FI}(\lambda) = (J\phi^{-1}(\lambda))^T \cdot FI(\phi^{-1}(\lambda)) \cdot J\phi^{-1}(\lambda).
\end{aligned}$$

Kaj je Jeffreysova apriorna porazdelitev na λ?

$$\begin{aligned}
&f_{Jeffrey}(\lambda) = c\sqrt{det\widetilde{FI}(\lambda)} = \\
&= c\sqrt{detFI(\phi^{-1}(\lambda))} = c|detJ(\phi^{-1}(\lambda))|;
\end{aligned}$$

to je transformirana Jeffreysova apriorna porazdelitev (transformirana sama vase).

Poglavje 3

Monte-Carlo integracija in metode vzorčenja

Proučujemo (neznani) parameter vzorčnega povprečja, recimo da nas zanima realnoštevilska funkcija $h(\theta)$.

Zgled.
Morda nas zanima „$E(X^2)$“ naše preučevane slučajne spremenljivke. Če imamo normalni model s parametrom $\theta = (\mu, \sigma^2)$, nas torej zanima $h(\mu, \sigma^2) = (\mu^2, \sigma^2)$. V Bayesovem okviru iz apriorne porazdelitve in vzorca dobimo aposteriorno porazdelitev z gostoto

$$f(\theta \mid x) = \frac{f(x \mid \theta) f(\theta)}{f(x)},$$

kot oceno za $h(\theta)$ vzamemo npr. aposteriorno pričakovano vrednost

$$E(h(\theta) \mid X = x) = \int h(\theta) f(\theta \mid x) d\theta.$$

Problemi:

- $h(\theta) f(\theta \mid x)$ je lahko zahtevna za integracijo - numerično,
- morda je „že“ $f(x \mid \theta) f(\theta)$ zahtevna za integracijo in $f(x)$ sploh ne „poznamo“.

Odgovor na to je integracija Monte-Carlo z Markovskimi verigami.

3.1 Klasična integracija Monte-Carlo

Zgled. „Integrirajte“ $\int_0^1 |\sin(100\sin(\pi x))|\,dx$.

Spomnimo se na KZVŠ: če so $X_1, X_2 \dots$ NEP slučajni vektorji s pričakovano vrednostjo μ, velja

$$\frac{x_1 + \cdots + x_n}{n} \to \mu = \int x f(x) dx$$

skoraj gotovo (s.g.).

Če je h taka realnoštevilska funkcija, da obstaja $E(h(X_i))$, so tudi $h(X_1), h(X_2) \dots$ NEP s pričakovano vrednostjo $E(h(X_i))$ in zato

$$\frac{h(x_1) + \cdots + h(x_n)}{n} \to E(h(X_i)) = \int h(x) f(x) dx$$

skoraj gotovo.

S pomočjo KZVŠ lahko ocenimo integral dane funkcije $h : (0{,}1) \to \mathbb{R}$ na naslednji način: privzamemo zaporedje NEP $X_i \sim N(0{,}1)$, torej velja

$$\frac{h(x_1) + \cdots + h(x_n)}{n} \to \int h(x).|.dx$$

skoraj gotovo.

Kaj to pomeni?

Verjetnost tistih zaporedij $(X_1, X_2 \dots)$, pri katerih limita $\frac{1}{n}\sum_{i=1}^{n} h(x_i)$ obstaja in je enaka .|..

Tu manjka ocena za natančnost ocene.

DN: implementirajmo.

NEP - izziv.

Izkaže se, da s psevdonaključnimi števili znamo izvrstno simulirati NEP. Vzorčenje iz (0,1) (za razumno velike vzorce).

Oceno natančnosti lahko dobimo s pomočjo CLI: privzemino obstoj disperzije slučajne spremenljivke X_i ($\iff \int h(x)^2 f(x) dx < \infty$).

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(h(X_i) - \overline{h(X_i)})^2.$$

Po CLI velja

$$\frac{h(X_i) - E(h(X_i))}{\frac{S}{\sqrt{n}}} \xrightarrow[n\to\infty]{D} N(0,1).$$

Za $\alpha \in (0, \frac{1}{2})$ sledi

$$P\left(\frac{\overline{h(X_i)} - E(h(X_i))}{\frac{S}{\sqrt{n}}} \in [-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]\right) = 1 - \alpha$$

Verjamemo, da se ocena $\overline{h(x_i)}$ razlikuje od dejanskega integrala $E(h(x_i))$ za kvečjemu $z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$.

3.2 Simulacija vzorčenja z inverzno kumulativno porazdelitveno funkcijo

Privzemimo, da je „ciljna“ kumulativna porazdelitvena funkcija (k.p.f.) $F : \mathbb{R} \to [0,1]$ zvezna bijekcija $\mathbb{R} \to (0,1)$. Velja trditev.

Trditev 3.2.1. Naj bo $U \sim U(0,1)$. Tedaj je $F^{-1}(U) \sim F$.

Posledica 3.2.2. Če „poznamo“ F^{-1}, znamo simulirati vzorčenje „iz F“.

Dokaz 3.2.3. $P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$.

Zgled. „Poznamo“ Φ^{-1}, kjer je $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

Zgornja trditev je standardna metoda za vzorčenje iz $N(0,1)$ ($\overset{\text{vaja}}{\Longrightarrow}$ znamo vzorčiti iz $N(\mu, \Sigma)$ za poljubne $\mu \in \mathbb{R}^d$ in $\sigma \in \mathbb{R}^{d\times d}$ s.p.d.).
V splošnem, če definiramo posplošeni inverz

$$F^{-}(U) = \inf F^{-1}([u, \infty)),$$

je (še vedno) $F^{-1}(U) \sim F$. ($\overset{\text{vaja}}{\Longrightarrow}$ znamo vzorčiti iz končnih diskretnih porazdelitev).

3.3 Metoda sprejmi ali zavrni (A/R)

Motivacija: Naj bo $f : \mathbb{R} \to [0, \infty)$ ciljna (Lebesgueova) gostota. Označimo „graf“ pod njo:

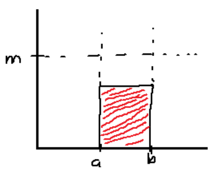
$$A_f = \{(x, u) \mid 0 \leq u \leq f(x)\} \subset \mathbb{R}^2.$$

Seveda je $S(A_f) = 1$ (ploščina).

Pripomnimo, da za $(X,U) \sim U(A_f)$ velja $X \sim f$:

$$f_X(x) = \int_{u=-\infty}^{\infty} f_{(X,U)}(x,u)du = \int_0^{f(x)} du = f(x).$$

A/R #1: privzemimo $\{x \mid f(x) > 0\} \subset (a,b)$, kjer $-\infty < a < b < \infty$ in $\exists m : f(x) < m$ za vse x.



Simulacijo vzorčenja $X \sim f$ implementiramo takole:

(i) vzorčimo $Y = y$, kjer $Y \sim (a,b)$ -„x“,

(ii) vzorčimo $(V \mid Y = y) = v$ (realizacija) iz $U(0,m)$ -„y“,

(iii) če je $v < f(y)$ sprejmemo $X = y$, če je $v \geq f(y)$ zavrnemo y in ponovimo (i).

Preverimo $X \sim f$. Za $I \subset (a,b)$ je

$$
\begin{aligned}
P(X \in I) &\overset{\text{vaja}}{=} P(Y \in I \mid v < f(y)) \\
&= \frac{\int_{\mathbb{R}} P(Y \in I \wedge v < f(Y) \mid Y = y) \cdot f_Y(y) dy}{\int_{\mathbb{R}} P(v < f(Y) \mid Y = y) \cdot f_Y(y) dy} \\
&\overset{\text{pogojna}}{=} \frac{\int_{\mathbb{R}} P(y \in I \wedge v < f(y)) \cdot f_Y(y) dy}{\int_{\mathbb{R}} P(v < f(y)) \cdot f_Y(y) dy} \\
&\overset{I \subset (a,b)}{=} \frac{\int_I P(v < f(y)) \cdot \frac{1}{b-a} dy}{\int_{(a,b)} P(v < f(y)) \cdot \frac{1}{b-a} dy} \\
&= \frac{\int_I f(y) dy}{\int_{(a,b)} f(y) dy} \\
&= \int_I f(y) dy,
\end{aligned}
$$

kjer smo v zadnjem koraku upoštevali $P(V < f(y)) = \frac{f(y)}{m}$ na (a,b). $\int_I f(y)dy$ je gostota X.

A/R #2: Privzemimo, da znamo vzorčiti iz gostote $g : \mathbb{R} \to [0,\infty)$ in da je $f(x) < M \cdot g(x)$ za vse x (za neki M). Simulacijo vzorčenja $X \sim f$ implementiramo takole:

(i) vzorčimo $Y = y$, kjer $Y \sim g$,

(ii) vzorčimo $(V \mid Y = y) = v$ iz $U(0,M \cdot g(y))$. Lahko vzamemo $W \sim U(0,1)$ in $V \sim M \cdot g(y) \cdot W$,

(iii) če je $v < f(y)$ sprejmemo $X = y$, če je $v \geq f(y)$ zavrnemo in ponovimo (i).

Preverimo $X \sim f$. Za $I \subset \mathbb{R}$ je

$$
\begin{aligned}
P(X \in I) &\overset{\text{vaja}}{=} P(Y \in I \mid M \cdot g(y) \cdot w < f(y)) \\
&= \frac{\int_{\mathbb{R}} P(Y \in I \wedge M \cdot g(y) \cdot w < f(Y) \mid Y = y) \cdot f_Y(y) dy}{\int_{\mathbb{R}} P(M \cdot g(y) \cdot w < f(Y) \mid Y = y) \cdot f_Y(y) dy} \\
&\overset{\text{pogojna}}{=} \frac{\int_I P(W < \frac{f(y)}{Mg(y)}) \cdot g(y) dy}{\int_{\mathbb{R}} P(W < \frac{f(y)}{Mg(y)}) \cdot g(y) dy} \\
&= \cdots = \int_I f(y) dy,
\end{aligned}
$$

kjer smo v upoštevali $P(W < \frac{f(y)}{Mg(y)}) = \frac{f(y)}{Mg(y)}$. Opazimo da je $M \cdot g(y)$ namesto m od prej (na nek način).
Pripomnimo, da je verjetnost sprejetja tu enaka

$$P(M \cdot g(y)W < f(y)) = \frac{1}{M}\int_{\mathbb{R}} f(y)dy = \frac{1}{M}.$$

Želimo M čim bližje 1.

Zgled. Oglejmo si $f = F_{N(0,1)}$ in $g = F_{Cauchy}$ $\left(g(x) = \frac{1}{\pi(1+x^2)}\right)$.

$$F_{Cauchy}(x) = \int_{-\infty}^{\infty} \frac{dt}{\pi(1+t^2)} = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$$
$$F_{Cauchy}^{-1}(u) = \tan\left(\pi\left(u - \frac{1}{2}\right)\right).$$

u smo izrazili iz $x = \frac{1}{\pi}arctan(x) + \frac{1}{2}$.
Vzorčenje iz Cauchyja je $\tan\big(\pi(U - \frac{1}{\pi})\big)$, U enakomerna na (0,1).

DN: optimiziraj M.

3.4 Metode MCMC (Monte Carlo z markovskimi verigami)

Okvir:
Želeli bi simulirati vzorčenje iz „ciljne“ spremenljivke z gostoto f (ki jo morda poznamo le do multiplikativne konstante natančno). Izkaže se, da za ocenjevanje pričakovane vrednosti

$$E(h(„X“)) = \int h(x)f(x)dx \quad \text{(nevtralne črke)}$$

pravzaprav ne potrebujemo simulacije neodvisnega vzorčenja.
Aproksimacija z vzročnim povprečjem

$$E_f(h) = \frac{1}{m}\sum_{i=1}^{m} h\left(X^{(i)}\right)$$

dobro funkcionira tudi v primeru, ko je $X^{(0)}, X^{(1)}, X^{(2)} \ldots$ primerna markovska veriga (z vrednostmi tam, kjer $f > 0$ - prostor „stanj“) s stacionarno porazdelitvijo z gostoto f.

Definicija 3.4.1 (Markovska veriga)**.** Markovska veriga je zaporedje s.v. (na prostoru stanj), ki ima lastnost, da je

(i) $\forall n$:

$$(X^{(n)} \mid X^{(n-1)} = x^{(n-1)} \ldots X^{(0)} = x^{(0)}) = (X^{(n)} \mid X^{(n-1)} = x^{(n-1)})$$

- markovska lastnost (neodvisno od n),

(ii) porazdelitve $(X^{(n)} \mid X^{(n-1)} = x)$ so za vse n enake (za vsak x imamo eno porazdelitev).

Tehnično gledano to pomeni, da Markovska veriga nastane tako, da ob času n vrednost $X^{(n)}$ dobimo z vzorčenjem iz cele porazdelitve, ki je odvisna le od stanja $x^{(n-1)}$.
$X^{(i)}$ - členi markovske verige,
$(X^{(n)} \mid X^{(n-1)} = x)$ - prehodne porazdelitve.
Markovska veriga (M.v.) ima stacionarno porazdelitev (...)f, če vedno velja sklep $(X^{(n-1)}) \sim f \implies X^{(n)} \sim f$.
(☢ nista pa $X^{(n-1)}$ in $X^{(n)}$ neodvisna s.v.)
Za ocenjevanje potrebujemo t.i. ergodične markovske verige:

(a) EZVŠ (ergodični zakon veliikih števil): če za vsako integrabilno funkcijo h velja

$$\lim_{n\to\infty} \frac{1}{n} h\left(X^{(i)}\right) = E_f(h)$$

skoraj gotovo (s.g.) za (skoraj) vse začetne vrednosti $x^{(0)}$ (ustrezne dobimo z verjetnostjo 1),

(b) ECLI (ergodični CLI): za vsako funkcijo h, za katero obstaja $\int h^2(x)f(x)dx$, obstaja konstanta γ_n (MCMC disperzija), za katero

$$\frac{\frac{1}{n}\sum_{i=1}^{n} h\left(X^{(i)}\right) - \int h(x)f(x)dx}{\frac{\gamma_n}{\sqrt{n}}} \xrightarrow{D} N(0{,}1).$$

(☢ γ_n je potrebno oceniti iz vzorca, kar je težko.)

Privzamemo (i) in naj bo $A \subset \{f > 0\}$ vsako območje, za katero

$$P(„X“ \in A) = \int_A f(x)dx = a > 0.$$

Vzenimo $a := 1_A$.
Tedaj $\frac{\text{število členov (do } n\text{-tega, ki padejo v)A}}{n} \overset{n\to\infty}{\to} \int 1_A(x)f(x)dx = a.$

3.4.1 Metropolisov algoritem

Naj bo f ciljna gostota in naj bo $\{q(y \mid x) \mid y,x \text{ iz prostora stanj}\}$ družina „predlaganih“ gostot: za vsak x (iz katerih pa znamo simulirati NEP vzorčenje) je $q(\_ \mid x)$ gostota neke porazdelitve. Naj bo še $q(y \mid x) = q(x \mid y)$ za vse pare (če smo v stanju x, predlagamo y „z enako verjetnsotjo“, kot če bi predlagali x, če smo v stanju y).
Opis algoritma:

(i) od nekod dobimo $x^{(0)}$,

(ii) privzamemo, da že imamo realizacijo $X^{(n-1)} = x^{(n-1)}$. Vzorčimo kandidata y za naslednjo realizacijo iz $q(\_ \mid x^{(n-1)})$.
Če velja $f(y) \geq f\left(x^{(n-1)}\right)$, vzamemo $X^{(n)} = y$ (realizacija).
Če je $f(y) < f\left(x^{(n-1)}\right)$, vzamemo

$$\begin{cases} X^{(n)} = y \text{ z verjetnostjo } \rho = \frac{f(y)}{f\left(x^{(n-1)}\right)} \\ X^{(n)} = x^{(n-1)} \text{ z verjetnostjo } 1-\rho \end{cases}$$

Naenkrat. $X^{(n)} = y$ z verjetnostjo $\rho = \min\{\frac{f(y)}{f\left(x^{(n-1)}\right)}, 1\}$ (*)
(*): če $f\left(x^{(n-1)}\right) = 0$, vedno vzamemo y.
Ta korak implementiramo z realizacijo $u \in U(0{,}1)$, vzamemo $X^{(n)} = y$, če $u \leq \rho$, oz. $X^{(n)} = x^{(n-1)}$, če $u > \rho$.

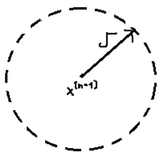
Izkaže se, da ta opis določa markovsko verigo s stacionarno porazdelitvijo f.
Če velja sklep $f(y) > 0 \implies \forall x : q(y \mid x) > 0$, ima veriga EZVŠ.
Bayesova aplikacija.
Ciljna gostota je $f(\theta \mid x)$(v θ).

Če je $\theta^{(n-1)}$ stanje v času $n-1$, za implementacijo koraka (ii) potrebujemo

$$\frac{f(\theta^* \mid x)}{f(\theta^{(n-1)} \mid x)} = \frac{f(x \mid \theta^*)f(\theta^{(*)})}{f(x \mid \theta^{(n-1)})f(\theta^{(n-1)})};$$

v resnici ne potrebujemo normalizacijske konstante v Bayesovi formuli.
Tipični primeri predlaganih gostot:

1. $(Y \mid X^{(n-1)} = x^{(n-1)}) \sim U\left(K_\delta(X^{(n-1)})\right)$ za fiksen δ (krogla s polmerom δ), v neki metriki.



   (Povrnljivost - povsod, kjer neničelne verjetnosti, jemlje ∞-krat.)
   (To pomeni $q(y \mid x) = \frac{1}{Vol(K_\delta(x))} \cdot 1_{K_\delta(x)}(y)$ (namesto klasične uporabimo ∞ metriko).)

2. $(Y \mid X^{(n-1)} = x^{(n-1)}) \sim N\left(x^{(n-1)}, \Sigma\right)$ za fiksno Σ.
   (To pomeni $q(y \mid x) = (2\pi)^{-\frac{dim}{2}}(det\Sigma)^{-\frac{1}{2}}e^{-\frac{1}{2}\langle\Sigma^{-1}(y-x),(y-x)\rangle}$ - simetričnost ✓.)
   V tem primeru je $Y|X^{(n-1)} = X^{(n-1)} + N(0, \Sigma)$.

Metropolisov algoritem tipično nima ECLI :(.

3.4.2 Metropolis-Hastingov algoritem

Tu predlagane gostote $q(y \mid x)$ ne zadoščajo simetričnosti. V algoritmu namesto ρ iz 3.4.1. uporabimo

$$\rho = \rho\left(x^{(n-1)}, y\right) = \min\{1, \frac{f(y)}{f\left(x^{(n-1)}\right)} \cdot \frac{q\left(x^{(n-1)} \mid y\right)}{q\left(y \mid x^{(n-1)}\right)}\}$$

(in $\rho = 1$ če $f\left(x^{(n-1)}\right) \cdot q\left(y \mid x^{(n-1)}\right) = 0$).
Enake lastnosti kot prej:

- vedno dobimo verigo s stacionarno porazdelitvijo f
- če $\forall x: \; q(\_, x^{(n)})$ dopušča kandidate iz $\{f > 0\}$ (v končno korakih), velja EVZŠ (blagi pogoji).

Dobimo pa še:
- pri primernih predpostavkah na q dobimo tudi ECLI.

Zgled („Neodvisni" Hastingov algoritem). Vedno „funkcionira" (teoretično) $q(y \mid x) = q(y)$ za neko fiksno porazdelitev z gostoto g, kjer $g(y) > 0$ za $\forall f(y) > 0$.

3.4.3 Gibsov vzorčevalnik

Gibsov vzorčevalnik je algoritem za konstrukcijo markovske verige s ciljno gostoto $f(x,y)$ (ali $f(x_1 \dots x_n)$) na podlagi vzorčenja iz „gostot" $f(x \mid y)$ ali $f(y \mid x)$.
Motivacija: proučujemo vzorčni model z gostotami $f(x \mid \theta_1, \theta_2) = f(x \mid \theta)$, ki je tak, da znamo simulirati neko vzorčenje iz $f(\theta_1 \mid \theta_2, x)$ in $f(\theta_2 \mid \theta_1, x)$, ne pa (neposredno) iz $f(\theta_1, \theta_2 \mid x)$.
Opis algoritma.

(i) Vzorčimo y_0 iz neke porazdelitve ali pa y_0 določimo.
Vzorčimo x_0 iz pogojne porazdelitve $f(X \mid y_0)$.

(ii) Če poznamo (x_{n-1}, y_{n-1}), vzorčimo najprej y_n iz $f(x \mid x_{n-1})$, potem pa še x_n iz $f(y \mid y_n)$ (temu koraku oz. njegovim podkorakom pravimo „osveževanje").

Dobimo zaporedje s.s. (ali DN)

$$X^{(0)} = (x_0, y_0),\; X^{(1)} = (x_1, y_1),\; X^{(2)} = (x_2, y_2) \text{ itd.}$$

Izkaže se, da so zaporedja

$$X^{(0)}, X^{(1)}, X^{(2)} \dots$$
$$X_0, X_1, X_2 \dots$$
$$Y_0, Y_1, Y_2 \dots$$

markovske verige in da ima veriga $\{X^{(i)} \mid n\}$ stacionarno porazdelitev $f(x,y)$. Pri blagih pogojih je ta veriga ergodična.

Zgled.
Tipična aplikacija v Bayesovi statistiki je:
privzemimo model $f(x \mid \theta_1, \theta_2)$ z apriorno gostoto $f(\theta_1, \theta_2) = f(\theta_1), f(\theta_2)$, kjer je $f(\theta_1)$ iz konjugirane družine k modelu $f(x \mid \theta_1, \text{KONST})$, $f(\theta_2)$ pa je iz konjugirane družine k modelu $f(x \mid \text{KONST}, \theta_2)$, iz katerih znamo simulirati NEP vzorčenje.
Polna aposteriorna porazdelitev

$$f(\theta_1, \theta_2 \mid x) = \frac{f(x \mid \theta_1, \theta_2) \cdot f(\theta_1) \cdot f(\theta_2)}{f(x)}$$

je tipično nedostopna, pač pa velja

$$f(\theta_1 \mid \theta_2, x) = \frac{f(x \mid \theta_1, \theta_2) \cdot f(\theta_1 \mid \theta_2)}{f(x \mid \theta_2)} = \frac{f(x \mid \theta_1, \theta_2) \cdot f(\theta_1)}{f(x \mid \theta_2)}, \tag{3.1}$$

kar je aposteriorna gostota Bayesovega modela z gostotami $f(x \mid \theta_1, \theta_2)$ iz apriorne gostote $f(\theta_1)$, kjer θ_2 razumemo kot konstanto. Po našem premisleku je torej $f(\theta_1 \mid \theta_2, x)$ iz „prave“ konjugirane družine. Simetrično je

$$f(\theta_2 \mid \theta_1, x) = \frac{f(x \mid \theta_1, \theta_2) \cdot f(\theta_2)}{f(x \mid \theta_1)}$$

iz „znane“ konjugirane družine.
Konkretno si oglejmo enorazsežni NEP-normalni model

$$(X \mid \mu, \sigma^2) \sim N\left(\begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \sigma^2 I\right) \tag{3.2}$$

z gostotami

$$f(x_1 \dots x_n \mid \mu, \sigma^2) = f(x \mid \mu, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (x_i - \mu)^2}.$$

Vemo, da je $\{N(\mu_*, \tau_*^2) \mid \mu_* \in \mathbb{R}, \tau_*^2 \in (0, \infty)\}$ konjugirana k enoparametričnim modelom z gostotami 3.2, kjer σ^2 poznamo.
Izkaže se (vaja), da je družina $\{\text{InvGama}(a,b) \mid a,b \in (0, \infty)\}$ konjugirana k enoparametričnim modelom z gostotami 3.2, kjer μ poznamo. Tu je $Y \sim \text{InvGama}(a,b) \iff \frac{1}{Y} \sim \text{Gama}(a,b)$, velja

$$f_{\text{InvGama}(a,b)}(y) = \frac{b^a}{\Gamma(a)} y^{-a-1} e^{-\frac{b}{y}}.$$

Vemo: pri $f(\mu) = f_{N(\mu_*, \tau_*^2)}(\mu)$ je

$$f(\mu \mid \sigma^2, x) \overset{3.1}{=} f_{N\left(\frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau_*^2}\mu_* + \frac{\tau_*^2}{\frac{\sigma^2}{n} + \tau_*^2}\overline{x}, \frac{\frac{\sigma^2}{n} \cdot \tau_*^2}{\frac{\sigma^2}{n} + \tau_*^2}\right)}(\mu).$$

Vidimo: pri $f(\sigma^2) = f_{\text{InvGama}(a,b)}(\sigma^2)$ je

$$f(\sigma^2 \mid \mu, x) = f_{\text{InvGama}\left(a + \frac{n}{2}, b + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right)}(\sigma^2).$$

<u>Gibsov vzorčevalnik</u>: ciljna porazdelitev $f(\mu, \sigma^2 \mid x)$.

(i) Določimo $\sigma_0^2 = 1$. Vzorčimo μ_0 iz

$$f_{N\left(\frac{\frac{1}{n}}{\frac{1}{n} + \tau_*^2}\mu + \frac{\tau_*^2}{\frac{1}{n} + \tau_*^2}\overline{x}, \frac{\frac{1}{n}\tau_*^2}{\frac{1}{n} + \tau_*^2}\right)},$$

(ii) vzorčimo σ_1^2 iz $f(\sigma^2 \mid \mu_0, x)$,
vzorčimo μ_1^2 iz ...
$\vdots$

(Blagi pogoji so izpolnjeni, veriga ergodična.)

3.5 Markovske verige z zveznim prostorom stanj - appendix

Markovska veriga „na“ Σ v diskretnem času je zaporedje merljivih preslikav $X_i : \Omega \to \Sigma$ $(i = 0, 1, 2 \dots)$, kjer je Ω verjetnostni prostor, Σ merljiv prostor in velja

$$P(X_n \in A \mid X_{n-1} = x_{n-1} \dots X_0 = x_0) = P(X_n \in A \mid X_{n-1} = x_{n-1}) \quad (3.3)$$

za vse $n \in \mathbb{N}, A \subset \Sigma$ in vse n-terice $x_0 \dots x_{n-1} \in \Sigma$.
Tu je $\{P(\_ \mid x) \mid x \in \Sigma\}$ družina verjetnostnih mer na Σ, za katero velja, da je za vsako merljivo množico $A \subset \Sigma$ preslikava $\Sigma \to [0{,}1]$, $x \mapsto \P(A \mid x)$ merljiva.
Prostoru Σ pravimo parameter stanj, X_i so členi verige, lastnosti 3.3 pa pravimo markovska lastnost.
Tu se bomo omejili na Borelove množice $\Sigma \subset \mathbb{R}^r$ (z Borelovo σ-algebro.)
Poudarimo, da so verjetnosti $P(_{|}x)$ (pravimo jim „prehodne verjetnosti“) neodvisne od n (torej za vse n enake).

Zgled.
Če ima X_0 neko porazdelitev in velja $X_n = X_{n-1} + \varepsilon_n$ (kjer so $X_0, \varepsilon_1 \dots$ neodvisne in $\epsilon_0 \sim N(0, V)$ za $V \in \P(r)$ (V je s.p.d.)) in $X_0, X_1, X_2 \dots$ markovska veriga in velja

$$P(A \mid x) = P(N(x, V) \in A).$$

Zgled. Metropolis-Hastingovova veriga:

(1) imamo $X_0 = x_0$,

(2) če je $X_{n-1} = x$, potem realizacijo od X_n dobimo takole:

   (i) vzorčimo y iz predlagane porazdelitve $dP_{(Y|x)} = q(y \mid x) d\nu(y)$

   (ii) vzorčimo $u = U \sim U(0{,}1)$ in sprejmemo $X_n = y$, če $u \leq \rho(x, y)$ oz. $X_n = x$, če $u > \rho(x{,}y)$, kjer

$$\rho(x{,}y) = \begin{cases} \min\{1, \frac{f(y)}{f(x)} \cdot \frac{q(x|y)}{q(y|x)}\}; f(x)q(y \mid x) \neq 0 \\ 1; f(x)q(y \mid x) = 0 \end{cases}$$

   Tu je f ciljna gostota (glede na ν),

(3)

$$
\begin{aligned}
&P(X_n \in A \mid X_{n-1} = x) \\
&= \int_{y\in\Sigma} P(X_n \in A \mid Y = y, X_{n-1} = x) dP_{(Y|x)}(x) \\
&= \int_{y\in\Sigma} \int_{u=0}^{1} P(X_n \in A \mid U = u, Y = y, X_{n-1} = x) \cdot dP_{(U|Y=y,X_{n-1}=x)}(u) d\nu(y) \\
&= \int_{y\in\Sigma} \left(\int_{u=0}^{\rho(x,y)} \ldots du + \int_{u=\rho(x,y)}^{1} \ldots du \right) q(y \mid x) d\nu(y) \\
&= \int_{y\in\Sigma} \left(1_A(y) \cdot q(y \mid x) + 1_A(x) \cdot (1 - \rho(x,y))\right) q(y \mid x) d\nu(y) \\
&= \int_{y\in A} \rho(x,y) q(y \mid x) d\nu(y) + 1_A(x) \int_{y\in\Sigma} (1 - \rho(x,y)) q(y \mid x) d\nu(y).
\end{aligned}
$$

Privzemimo, da ν nima atomov: $\nu(x) = 0$ za vsak x:

$$
\implies P(X_n \in A \mid X_{n-1} = x) = \int_{y\in\Sigma} (1 - \rho(x,y)) q(y \mid x) d\nu(y).
$$

To je tipično u pozitivna verjetnost. To pomeni, da imajo prehodne verjetnosti atome u: $P(x \mid x) > 0$ (vsaj za nekatere x).

Definicija 3.5.1.
Porazdelitev π je stacionarna za verigo $\{X_i\}$, če iz $X_{n-1} \sim \pi$ sledi $X_n \sim \pi$.

Trditev 3.5.2.
Porazdelitev $f d\nu$ je stacionarna porazdelitev M-H verige.

Dokaz 3.5.3.
Privzemimo $dP_{X_{n-1}} = f d\nu$ in računajmo

$$
\begin{aligned}
P(X_n \in A) &= \int_{x\in\Sigma} P(X_n \in A \mid X_{n-1} = x) dP_{X_{n-1}}(x) \\
&= \int_{x\in\Sigma} \int_{y\in A} \rho(x,y) q(y \mid x) f(y) d\nu(y) d\nu(x) \\
&\quad + \int_{x\in A} \int_{y\in\Sigma} q(x \mid y) d\nu(y) f(x) d\nu(x) \\
&\quad - \int_{x\in A} \int_{y\in\Sigma} \rho(x,y) q(x \mid y) f(x) d\nu(y) d\nu(x).
\end{aligned}
$$

Upoštevamo

$$
\int_{y\in\Sigma} q(x \mid y) f(x) d\nu(y) = 1.
$$

Spomnimo se:

$$\rho(x,y) = \begin{cases} \min\{1, \frac{f(y)}{f(x)} \cdot \frac{q(x|y)}{q(y|x)}\}; f(x)q(y \mid x) \neq 0 \\ 1; f(x)q(y \mid x) = 0 \end{cases}$$

Vidimo, da velja enakost (za $\forall x, y$)

$$\rho(x,y)q(y \mid x)f(x) = \rho(y,x)q(x \mid y)f(y)$$

in

$$\int_{x\in\Sigma}\int_{y\in A} \rho(x,y)q(y \mid x)f(y)d\nu(y)d\nu(x) = \int_{y\in\Sigma}\int_{x\in A} \rho(y,x)q(x \mid y)f(x)d\nu(x)d\nu(y).$$

Zato je

$$P(X_n \in A) = \int_{x\in A} f(x)d\nu(y).$$

Prehodne verjetnosti $P(_| x)$ generirajo verigo prehodnih verjetnosti $P(A \mid x) = P(X_n \in A \mid X_0 = x)$.

To verjetnost, da veriga obišče A v času n ob pogoju, da je začela v x, željeni lastnosti (ki implicirata ustrezna ERVŠ) sta

- $||P^n(_| x) - \pi|| \overset{n\to\infty}{\to} 0$ za skoraj vse x,
- $||P^n(_| x) - \pi|| \overset{n\to\infty}{\to} 0$ enakomerno v x,

kjer za predznačeno mero μ definiramo normo totalne variacije

$$||u|| = \sup_A |\mu(A)|.$$

Za osnovne lastnosti $P^n(A \mid x)$ izračunajmo

$$\begin{aligned}
&P(X_n \in A_n, X_{n-1} \in A_{n-1} \ldots X_1 \in A_1 \mid X_0 = x_0) \\
&= \int_{x_1\in\sigma} P(X_n \in A_n \ldots X_1 \in A_1 \mid X_1 = x_1, X_0 = x_0) dP_{(X_1|X_0=x_0)}(x_1) \\
&= \int_{x_1\in A_1} P(X_n \in A_n \ldots X_2 \in A_2 \mid X_1 = x_1, X_0 = x_0) dP_{(X_1|X_0=x_0)}(x_1 \mid x_0) \\
&= \int_{x_1\in A_1}\int_{x_2\in\Sigma} P(X_n \in A_n \ldots X_2 \in A_2 \mid X_2 = x_2, X_1 = x_1, X_0 = x_0) dP(x_2 \mid x_1) dP(x_1 \mid x_0) \\
&= \ldots \\
&= \int_{x_1\in A_1} \cdots \int_{x_n\in A_n} dP(x_n \mid x_{n-1}) \ldots dP(x_1 \mid x_0).
\end{aligned}$$

Posebej:

$$P^n(A \mid x) = \int_{x_1\in\Sigma} \cdots \int_{x_{n-1}\in\Sigma} \int_{x_n \in A} dP(x_n \mid x_{n-1}) \ldots dP(x_1 \mid x).$$

Od tod sledi

$$P^n(A \mid x) = \int_{x_1\in\Sigma} \cdots \int_{x_m\in\Sigma} \left(\int_{x_{m+1}\in\Sigma} \cdots \int_{x_n\in A} dP(x_n \mid x_{n-1} \ldots dP(x_{m+1} \mid x_m)) \right) \\ \cdot dP(x_m \mid x_{m-1}) \ldots dP(x_1 \mid x_0).$$

Tukaj je

$$\int_{x_{m+1}\in\Sigma} \cdots \int_{x_n\in A} dP(x_n \mid x_{n-1} \ldots dP(x_{m+1} \mid x_m)) = P^{n-m}(A \mid x_m)$$

in zato

$$P^n(A \mid x) = \int_{x_1\in\Sigma} P^{n-m}(A \mid x_m) \cdot dP^m(x_m \mid x).$$

Tej enakosti rečemo enakost Chapman-Kolmogorova.

3.6 MCMC diagnostika

3.6.1 MCMC varianca

Želimo oceniti $E_f(h) = \int h(x)f(x)dx$.

Privzemimo, da je $X_0, X_1 \ldots$ markovska veriga s stacionarno porazdelitvijo „f“, ki ima primerne ergodične lastnosti.

Označimo $\widehat{E_f(h)} = \frac{1}{n}\sum_{i=1}^n h(X_i)$ (standardna MCMC cenilka za $E_f(h)$ za veriga do časa n).

Definirajmo $D_{MCMC}\left(\widehat{E_f(h)(x)}\right) = E\left(\left(\widehat{E_f(h)(x)} - E_f(h)\right)^2\right)$.

(To je v resnici SNK glede na ocenjeno karakteristiko $E_f(h)$. V splošnem $\widehat{E_f(h)}$ ni nepristranska cenilka za $E_f(h)$.)

Tu je $\widehat{E_f(h)} - E_f(h) = \frac{1}{n}\sum_{i=1}^n (h(X_i) - E_f(h))$, velja

$$D_{MCMC}\left(\widehat{E_f(h)}\right) = \frac{1}{n^2}\sum_{i=1}^{k} E((h(X_i) - E_f(h))^2) + \frac{1}{n^2}\sum_{j\neq j} E((h(X_i) - E_f(h))(h(X_j) - E_f(h))).$$

Privzemimo, da je veriga STACIONARNA, t.j. $X_i \sim f$ za vse i (to sledi iz $X_0 \sim f$). Tedaj je

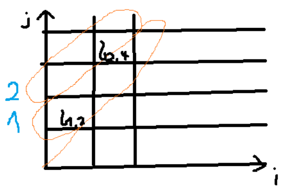
(i) $E((h(X_i) - E_f(h))^2) =: \sigma^2$ varianca („s.s.“ $h(X_i)$) (enaka za vse i) in

(ii) $\sigma_{i,j} = E((h(X_i) - E_f(h))(h(X_j) - E_f(h)))$ kovarianca „s.s.“ $h(X_i)$ in $h(X_j)$, ki je odvisna le od $|i-j|$: če je npr $i < j$, je

$$E(h(X_i)h(X_j)) = \int E(h(x_i)h(x_j) \mid X_{i-1} = x_{i-1})dx_{i-1}.$$

Sledi $D_{MCMC}\left(\widehat{E_f(h)}\right) = \frac{\sigma^2}{n} + \frac{2}{n^2}\sum_{i<j}\sigma_{i,j}$ oziroma

$$\begin{aligned} nD_{MCMC}\left(\widehat{E_f(h)}\right) &= \sigma^2 + \frac{2}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sigma_{i,j} \\ &= \sigma^2 + \frac{2}{n}\sum_{k=1}^{n-1}(n-k)\sigma_{0,k} \\ &= \sigma^2 + 2\sum_{k=1}^{n-1}\sigma_{0,k} - \frac{2}{n}\sum_{k=1}^{n-1}k\sigma_{0,k}. \end{aligned}$$



Izkaže se, da pri primernih ergodičnih lastnostih velja $\sum_{k=1}^{\infty}|\sigma_{0,k}| < \infty$ in posledično $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n-1}k\sigma_{0,k} = 0$.
Zato definiramo asimptotično MCMC varianco kot

$$\gamma_k^2 = \sigma^2 + 2\sum_{k=1}^{\infty}\sigma_{0,k},$$

kjer je σ^2 „stacionarna“ varianca, $\sigma_{0,k}$ pa so „stacionarne rang-k“ avtokorelacijske verige.

Izrek 3.6.1. Pri primernih ergodičnih lastnostih velja

$$\text{(ECLI)}: \quad \frac{\widehat{E_f(h)} - E_f(h)}{\frac{\gamma_k}{\sqrt{n}}} \longrightarrow N(0,1)$$

(za KATEROKOLI začetno porazdelitev).

Je to že dovolj za konstrukcijo (asimptotičnega) intervala zaupanja za $E_f(h)$? Teoretično da:

$$\lim_{n\to\infty} P\left(E_f(h) \in \left[\widehat{E_f(h)} - z_{\frac{\alpha}{2}} \cdot \frac{\gamma_k}{\sqrt{n}}, \widehat{E_f(h)} + z_{\frac{\alpha}{2}} \cdot \frac{\gamma_k}{\sqrt{n}}\right]\right) = 1 - \alpha.$$

Za praktično rabo je treba γ_k^2 oceniti iz vzorca. Preprosto je oceniti σ^2; vzamemo kar

$$\frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(h(X_i) - E_f(X_i)\right)^2.$$

Vsoto vrste $\sum_{k=1}^{n} \sigma_{0,k}$ je zahtevno oceniti.

Za neposredno ocenjevanje γ_k^2 si oglejmo metodo povprečij neprikrivajočih se serij.

Privzemimo, da velja $n = ab$ za neki $a, b \in \mathbb{N}$. Veriga $X_1 \dots X_n = X_{ab}$ razdelimo v serije („batches“)

$X_1 \dots X_b \overset{\text{povprečimo}}{\longrightarrow} \widehat{\mu_{b,1}}$

$X_{b+1} \dots X_{2b} \longrightarrow \widehat{\mu_{b,2}}$

$\dots$

$X_{(a-1)b+1} \dots X_{ab} \longrightarrow \widehat{\mu_{b,a}}$.

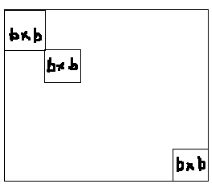
Pripomnimo, da je $\widehat{\mu_{b,k}} = \frac{1}{b} \sum_{i=(k-1)b+1}^{kb} h(X_i)$.

Tudi to je MCMC ocena za $E_f(h)$. Pišimo še $\widehat{\mu_n} = \widehat{\mu_{ab}} = \frac{1}{a} \sum_{i=1}^{n} h(X_i)$.

Potem je $\widehat{\gamma_k^2} = \frac{b}{a} \sum_{k=1}^{a} (\widehat{\mu_{b,k}} - \widehat{\mu_k})^2$ ocena za γ_k^2.

Pišimo $\mu = E_f(h)$. Če začnemo z $\widehat{\gamma_k^2} = \frac{b}{a} \sum_{k=1}^{a} (\widehat{\mu_{b,k}} - \mu + \mu - \widehat{\mu_b})^2$, račun pripelje

$$\widehat{\gamma_k^2} = \frac{1}{n} \sum_{i=1}^{n} (h(X_i) - \mu)^2 + \frac{1}{a} \sum_{k=1}^{a} \frac{1}{b} \sum_{i=(k-1)b+1, i\neq j}^{kb} ((h(X_i) - \mu)(h(X_j) - \mu))$$
$$- \frac{1}{an} \left(\sum_{i=1}^{n} (h(X_i) - \mu)^2 + \sum_{i\neq j} (h(X_i) - \mu)(h(X_j) - \mu) \right).$$



$\frac{1}{n} \sum_{i=1}^{n} (h(X_i) - \mu)^2 \overset{n\to\infty}{\longrightarrow} \sigma^2$ (EZVŠ),

$\frac{1}{a}\sum_{k=1}^{a}\frac{1}{b}\sum_{i=(k-1)b+1,i\neq j}^{kb}((h(X_i)-\mu)(h(X_j)-\mu)) \overset{a\to\infty,\, b\to\infty}{\longrightarrow} 2\sum_{k=1}^{\infty}\sigma_{0,k}$,
ostalo $\overset{a\to\infty,\, b\to\infty}{\longrightarrow} 0$.

Priporočeni izbiri sta $a=\sqrt{n}$, $a=\sqrt[3]{n}$.

V praksi to realiziramo kot $a(k)\cdot b(k)=n(k)=k\cdot k^2$ $k=1,2\ldots$

Sledi želeni ECLI:

$\lim_{n\to\infty}\frac{\widehat{E_f(h)}n(k)-E_f(h)}{\sqrt{\widehat{\gamma_k^2}(a(k),b(k))/\sqrt{n}}} \overset{D}{=} N(0{,}1)$

in še EIZ (ergodični interval zaupanja) (stopnje zaupanje $1-\alpha$)

$\left[\widehat{E_f(h)}-z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{\gamma_k^2}(a(k),b(k))}{a(k)\cdot b(k)}}, \widehat{E_f(h)}+z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{\gamma_k^2}(a(k),b(k))}{a(k)\cdot b(k)}}\right]$.

Efektivna velikost vzorca.

Naj bo $\sigma^2=\int(h(x)-E_f(x))^2 f(x)dx$ kot prej.

Če bi znali vzorčiti (NEP) iz f, bi vzorčno povprečje imelo varianco $\frac{\sigma^2}{n}$.

Za MCMC vzorčenje imamo $D_{MCMC}(\widehat{E_f(h)})\approx\frac{\gamma_k^2}{n}$.

Definicija 3.6.2. Efektivna velikost vzorca n_{eff} za dejansko velikost vzorca n je rešitev enačbe $\frac{\gamma_k^2}{n}=\frac{\sigma^2}{n_{eff}}$.

Pri danem vzorcu seveda n_{eff} ocenimo z

$$\widehat{n_{eff}}=n\frac{\widehat{\sigma^2}}{\gamma_k^2}=\frac{\widehat{\sigma^2}}{\gamma_k^2/n}.$$

3.6.2 Mešanje

Kvaliteto konvergence dane markovske verige za dano velikost vzorca n lahko ocenimo tudi z „mešanjem“ neodvisnih verig, ki začnejo v „mešano“ različnih točkah prostora stanj.

Privzemimo torej, da so X_{ij} $1\leq i\leq n$ neodvisne markovske verige z začetkom v $X_{01}\ldots X_{0m}$ (torej $1\leq j\leq m$).

Tu so X_{0j} $(1\leq j\leq m)$ lahko konstruirani „ročno“ ali pa jih vzorčimo iz neke fiksne porazdelitve z veliko disperzijo.

Povzemimo: skupine $(X_{1j}\ldots X_{nj})$ so med seboj neodvisne $(1\leq j\leq m)$, $X_{i-1,j}\to X_{i,j}$ pa so dani z prehodno verjetnostjo.

<u>Označimo</u>

$W=\frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(h(X_{ij})-\overline{h(X_{.j})}\right)^2$ (variance „znotraj“ skupin),

$B = \frac{n}{m-1} \sum_{j=1}^{m} \left(\overline{h(X_{.j})} - \overline{h(X_{..})} \right)^2$ (varianca „med“ skupinami).
Končno definiramo še $\widehat{\sigma^{2+}} = \widehat{D_f(h)^+} (= \frac{n}{n-1} W + \frac{1}{n} B$.
Izkaže se, da $\widehat{\sigma^{2+}}$ precenjuje σ^2, medtem ko W podcenjuje σ^2.
Velja $\lim_{n\to\infty} \sqrt{\frac{\widehat{\sigma^{2+}}}{W}} = 1$;
če imamo za dami k $\frac{\widehat{\sigma^{2+}}}{W} >> 1, n$ povečujemo.

Poglavje 4

Normalni modeli

Uvodni zgled: 1-razsežni „NEP-normalni“ model, kjer je $(X \mid \mu, \sigma^2) \sim N(\left(\begin{bmatrix} \mu & \vdots & \mu \end{bmatrix}, \sigma^2 I_{n\times n}\right))$ z vzorčnimi gostotami

$$f(x \mid \mu, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i-\mu)^2\right)}. \tag{4.1}$$

Izkaže se, da je konjugirana družina apriornih gostot podana kot

$$f(\mu, \sigma^2) = f_{\mathrm{InvGama}(a,b)}(\sigma^2) \cdot f_{N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)}(\mu). \tag{4.2}$$

(Za vajo lahko izpeljete; razcepite $\psi = \psi_1 + \psi_2$ (log ?) $\to \tau_1, \tau_2$ namesto τ). Tu so $\mu_0 \in \mathbb{R}, a, b, \kappa_0 \in (0, \infty)$ parametri konjugirane družine; κ_0 interpretiramo kot število prostorskih stopenj (fiktivno).

4.1 Dvofazna predstavitev

Privzemimo razcep $\vartheta = (\vartheta_1, \vartheta_2)$ in zapišimo

$$f(\vartheta_1, \vartheta_2) = f(\vartheta_1 \mid \vartheta_2) \cdot f(\vartheta_2).$$

Za aposteriorno gostoto pri $X = x$ velja

$$f(\vartheta_1, \vartheta_2 \mid x) = \frac{f(x \mid \vartheta_1, \vartheta_2) \cdot f(\vartheta_1 \mid \vartheta_2)}{f(x \mid \vartheta_2)} \cdot \frac{f(x \mid \vartheta_2) \cdot f(\vartheta_2)}{f(x)}$$

oz.

$$f(\vartheta_1, \vartheta_2 \mid x) = f(\vartheta_1 \mid \vartheta_2, x) \cdot f(\vartheta_2 \mid x).$$

Tu je $f(\vartheta_1 \mid \vartheta_2, x)$ aposteriorna gostota modela z vzorčnimi gostotami $f(x \mid \vartheta_2)$ in apriorno gostoto $f(\vartheta_2)$.
Tipično je $f(\vartheta_1 \mid \vartheta_2, x)$ enostavneje izračunati (eksplicitno) kot $f(\vartheta_2 \mid x)$, ker za slednje potrebujemo še $f(x \mid \vartheta_2)$.
Načeloma lahko $f(x \mid \vartheta_2)$ izračunamo z integriranjem:

$$f(x \mid \vartheta_2) = \int f(x \mid \vartheta_1, \vartheta_2) \cdot f(\vartheta_1 \mid \vartheta_2) d\vartheta_1,$$

vendar :(
Aplicirajmo na vzorčne gostote z 4.1 z apriorno 4.2.

(i) $f(\mu \mid \sigma^2, x)$ „pripada“ modelu $(X \mid \mu) \sim N\left(\begin{bmatrix}\mu \\ \vdots \\ \mu\end{bmatrix}, \sigma^2 I\right)$ (σ^2 konst.)
za apriorno $M \sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) = N(\mu_0, \tau_0^2)$.
To pomeni $f(\mu \mid \sigma^2, x) \leftrightarrow N(\mu, \tau_0^2)$, kjer je

$$\mu_1 = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau_0^2}\mu_0 + \frac{\tau_0^2}{\frac{\sigma^2}{n} + \tau_0^2}\overline{x}$$

in

$$\tau_1^2 = \frac{\frac{\sigma^2}{n} \cdot \tau_0^2}{\frac{\sigma^2}{n} + \tau_0^2} = \frac{\sigma^2}{n + \kappa_0}.$$

Interpretacija variance: več primerov.

(ii) Za $f(\sigma^2 \mid x)$ potrebujemo $f(x \mid \sigma^2)$.
DN: intergirajte.

Oglejmo si raje

$$
\begin{aligned}
f(x, \mu \mid \sigma^2) &= f(x \mid \mu, \sigma^2) \cdot f(\mu \mid \sigma^2) \\
&\propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \cdot e^{-\frac{1}{2\left(\frac{\sigma^2}{\kappa_0}\right)}(\mu-\mu_0)^2} \\
&= e^{-\frac{1}{2}\left\langle W\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ \mu \end{bmatrix}, \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ \mu \end{bmatrix}\right\rangle},
\end{aligned}
$$

kjer je

$$
W = \begin{bmatrix} \frac{1}{\sigma^2} & \dots & 0 & -\frac{1}{\sigma^2} \\ 0 & \dots & \vdots & -\frac{1}{\sigma^2} \\ \vdots & \vdots & \frac{1}{\sigma^2} & -\frac{1}{\sigma^2} \\ -\frac{1}{\sigma^2} & \frac{1}{\sigma^2} & \dots & \frac{\kappa_0+n}{\sigma^2} \end{bmatrix}.
$$

Vidimo, da je $(X, M \mid V \sim \sigma^2)$ normalna porazdelitev $\implies$ $(X \mid V \sim \sigma^2)$ je kot robna tudi normalna.
$\overset{!!}{\implies}$ potrebujemo le $E(X \mid V \sim \sigma^2)$ in $Var(X \mid V \sim \sigma^2)$.
Vemo:

$$
\begin{aligned}
E(X \mid V \sim \sigma^2) =& E(E(X \mid M, V) \mid V \sim \sigma^2) \text{ in} \\
Var(X \mid V \sim \sigma^2) =& E(Var(X \mid M, V) \mid V \sim \sigma^2) + \\
& Var(E(X \mid M, V) \mid V \sim \sigma^2).
\end{aligned}
$$

Naprej:

$$
E(X \mid M = \mu, V \sim \sigma^2) = \mu \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \implies E(X \mid M, V) = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \cdot M
$$

$$
Var(X \mid M = \mu, V \sim \sigma^2) = \sigma^2 I \implies Var(X \mid M, V) = V \cdot I.
$$

Sledi:

$$E(X \mid V \sim \sigma^2) = E\left(\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} M \mid V \sim \sigma\right) = \begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}\mu_0 \text{ in}$$

$$Var(X \mid V \sim \sigma^2) = \sigma^2 I + \begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} Var(M \mid V \sim \sigma^2)\begin{bmatrix}1 & \cdots & 1\end{bmatrix}$$

$$= \sigma^2\left(I + \frac{1}{\kappa_0}\begin{bmatrix}1 & \dots & 1\\ \vdots & & \vdots\\ 1 & \dots & 1\end{bmatrix}\right).$$

Lahko zapišemo

$$f(x \mid \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot det\left(I + \frac{1}{\kappa_0}\begin{bmatrix}1 & \dots & 1\\ \vdots & & \vdots\\ 1 & \dots & 1\end{bmatrix}\right)^{-1}$$

$$\cdot e^{-\frac{1}{2\sigma^2}\left\langle\left(I+\frac{1}{\kappa_0}\begin{bmatrix}1 & \dots & 1\\ \vdots & & \vdots\\ 1 & \dots & 1\end{bmatrix}\right)^{-1}\left(x-\begin{bmatrix}\mu_0\\ \vdots\\ \mu_0\end{bmatrix}\right), x-\begin{bmatrix}\mu_0\\ \vdots\\ \mu_0\\ \mu_0\end{bmatrix}\right\rangle}.$$

Velja še

$$f(\sigma^2) = f_{\text{InvGama}}(\sigma^2) = \frac{b^a}{\gamma(a)} \cdot (\sigma^2)^{-a-1} \cdot e^{-\frac{b}{\sigma^2}}$$

$$\implies f(\sigma^2 \mid x) \propto (\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{1}{\sigma^2}\frac{1}{2}(\langle \dots, \dots \rangle)} \cdot (\sigma^2)^{-a-1} e^{-\frac{b}{\sigma^2}}$$

$$\leftrightarrow \text{InvGama}\left(a + \frac{n}{2}, b + \frac{1}{2}\left\langle\left(I + \frac{1}{\kappa_0}\begin{bmatrix}1 & \dots & 1\\ \vdots & & \vdots\\ 1 & \dots & 1\end{bmatrix}\right)^{-1}\left(x - \begin{bmatrix}\mu_0\\ \vdots\\ \mu_0\end{bmatrix}\right), x - \begin{bmatrix}\mu_0\\ \vdots\\ \mu_0\end{bmatrix}\right\rangle\right.$$

$\leftrightarrow$: konjugirana porazdelitev.

Prepričali smo se, da je opisana družina apriornih porazdelitev konjugirana, in sicer

- $a \to a + \frac{n}{2}$
- $b \to b + \frac{1}{2}\langle \ldots, \ldots \rangle$
- $\mu_0 \to \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\overline{x}$
- $\kappa_0 \to \kappa_0 + n$.

Nova b in μ_0 sta odvisna od realizacije $(x_1 \ldots x_n)$.
$b + \frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{n\cdot\kappa_0}{\kappa_0+n}(\overline{x} - \mu_0)^2\right)$

4.2 Hierarhični modeli

Spomnimo se na normalni model „preizkušnja m terapij". Frekvencistično gledamo neodvisne skupine s.s.

$$\begin{array}{cccc} X_{1,1} & X_{1,2} & \ldots & X_{1,m} \\ \vdots & \vdots & & X_{n_m,m} \\ X_{n_1,1} & \vdots & & \\ & X_{n_2,2} & & \end{array}.$$

Velikosti $n_1 + \cdots + n_m = n$.
Funkcija neodvisnosti: $X_{ij} \sim N(\mu_j, \sigma^2)$.
Vse variance so enaka (homoskelastičnost) zaradi „preprostosti".
Najbolj nas zanimajo (ocene) za μ_j in σ^2, „optimalne" frekventistične cenilke so:

$$\hat{\mu_j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij} = \overline{X_{.j}}$$
$$\hat{\sigma} = \frac{1}{n-m}\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij} - \overline{X_{.j}})^2.$$

Zadnjič: model m terapij $N(\mu_j, \sigma^2)$ $(1 \leq j \leq m)$.
Po preizkušnju domnev je glavna t.i. domneva homogenosti
$\mu_1 = \mu_2 = \cdots = \mu_m$.
V pridruženem Bayesovem modelu začnemo z idejo, da so μ_j „poustvarjanje" nekega „izhodiščnega" μ_0.

σ^2 našega homoskelastičnega modela je znana.
μ_0, τ_0^2, a, b
$\downarrow$
$\mu, \eta^2 \in N(\mu_0, \tau_0^2) \times$ InvGama(a,b) - Bayesov parameter
$\downarrow$
$\mu_1 \dots \mu_m \in N(\mu, \eta^2) \times \cdots \times N(\mu, \eta^2)$ - ϕ - Bayesov parameter
$\downarrow$ (neodvisno)
$x_{i_1} \in N(\mu_1, \sigma^2)$ $(1 \leq i \leq n_1) \dots x_{i_m} \in N(\mu_m, \sigma^2)$ $(1 \leq i \leq n_m)$ - ϑ - vzorec (večji τ^2, bolj μ_i različni).

4.2.1 Abstraktna opredelitev hierarhičnega modela

Bayesov parameter je oblike (ϑ, ϕ) kjer ϕ imenujemo hiperparameter, ϑ pa populacijski parameter. Za vzorčne gostote velja temeljni privzetek
(HM - hierarhični model) $f(x \mid \vartheta, \phi) = f(x \mid \vartheta)$.
Aposteriorne gostote bi lahko predstavili kot

$$\begin{aligned} f(\vartheta \mid \phi, x) &= \frac{f(x \mid \vartheta, \phi) f(\vartheta \mid \phi) f(\phi)}{f(x)} \\ &= \frac{f(x \mid \vartheta) f(\vartheta)}{f(x)} \cdot \frac{f(\vartheta \mid \phi) f(\phi)}{f(\vartheta)} \\ &= f(\vartheta \mid x) f(\phi \mid \vartheta). \end{aligned}$$

Ta predstavitev tipično ni uporabna, ker ne poznamo $f(\vartheta)$. Zato raje

- $f(\vartheta \mid \phi, x) = \frac{f(x|\vartheta,\phi) f(\vartheta|\phi)}{f(x|\phi)}$: aposteriorna gostota modela z vzorčnimi $f(x \mid \vartheta)$ in apriornimi $f(\vartheta \mid \phi)$,
- $f(\phi \mid \vartheta, x) = \frac{f(x|\vartheta) f(\phi|\vartheta)}{f(x|\vartheta)}$: aposteriorna gostota modela z vzorčnimi $f(\vartheta \mid \phi)$ in apriornimi $f(\phi)$

($\implies$ z Gibsom $f(\vartheta, \phi \mid x)$).
Aplicirajmo na naš zgled:
$\phi = (\mu, \eta^2)$

$n = \sum_{j=1}^{m} n_j$
$\vartheta = (\mu_1 \dots \mu_m)$

$$f([x_{ij}] \mid \mu_1 \dots \mu_m) = \prod_{j=1}^{m} \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_j}(x_{ij}-\mu_j)^2}$$
$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}(x_{ij}-\mu_j)^2}$$

$j = 1 \dots m,\ i = 1 \dots n_j.$
Populacijska apriorna:

$$f(\mu_1 \dots \mu_m \mid \mu, \eta^2) = \left(2\pi\eta^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\eta^2}\sum_{j=1}^{m}(\mu_j-\mu)^2},$$

hiperapriorna:

$$f(\mu, \sigma^2) = f_{N(\mu_0,\tau_0^2)}(\mu) \cdot f_{\text{InvGama}(a,b)}(\eta^2).$$

$$f(\mu_1 \dots \mu_m \mid \mu, \eta^2, [x_{ij}])$$
$$\propto \prod_{i=1}^{m} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_j}(x_{ij}-\mu_j)^2} \cdot e^{-\frac{1}{2\eta^2}(\mu_j-\mu)^2}.$$

Analogije z aposteriornim NEP-normalnim modelom:

- $\sigma^2 - \sigma^2$
- $x_{ij} - x_i$
- $\mu_j - \mu$
- $\eta^2 - \tau_0^2$
- $\mu_j - \mu$
- $\mu - \mu_0$.

Ta pogojna aposteriorna porazdelitev je

$$\prod_{j=1}^{m} N\left(\frac{\frac{\sigma^2}{n_j}}{\frac{\sigma^2}{n_j}+\eta^2}\mu + \frac{\eta^2}{\frac{\sigma^2}{n_j}+\eta^2}\overline{x_{\cdot j}}, \frac{\frac{\sigma^2}{n_j}\cdot\eta^2}{\frac{\sigma^2}{n_j}+\eta^2}\right).$$

$f(\mu, \eta^2 \mid \mu_1 \dots \mu_m)$ 1-razsežni NEP-normalni model s polkonjugirano apriorno porazdelitvijo

$$f(\mu \mid \eta^2, \mu_1 \dots \mu_m) \dots N\left(\frac{\frac{\tau_0^2}{m}}{\frac{\tau_0^2}{m}+\tau_0^2}\mu_0 + \frac{\tau_0^2}{\frac{\tau_0^2}{m}+\tau_0^2}\overline{\mu}, \frac{\frac{\tau_0^2}{m}\cdot\tau_0^2}{\frac{\tau_0^2}{m}+\tau_0^2}\right)$$
$$f(\tau^2 \mid \mu, \mu_1 \dots \mu_m) \dots \mathrm{InvGama}\left(a + \frac{m}{2}, b + \frac{1}{2}\sum_{i=1}^{m}(\mu_j - \mu)^2\right).$$

Definicija 4.2.1. Naj bo ν mera na $B(\Sigma)$. Tedaj je veriga $\{X_n\}$ ν-ireducibilna, če iz $\nu(A) > 0$ sledi

$$\forall x\ \exists m = n_n:\ P^n(A \mid x) = P(X_m \in A \mid X_0 = x) > 0.$$

Če naj velja EZVŠ, moramo vsako množico A, za katero

$$\pi(A) = \int_A d\pi(y) = \int_A f(y)dy > 0$$

obiskati neskončno-mnogokrat. Torej nas v našem kontekstu zanima π-ireducibilnost, kjer je π stacionarna porazdelitev.

Primer. Če iz $f(y) > 0$ sledi $q(y \mid x) > 0$, je M-H (Metropolis-Hastingova) veriga π-ireducibilna.

Komentar: zgornji pogoj je za praktične namene pogosto pomešan? (neuporaben?).

Definicija 4.2.2. Naj bo π stacionarna porazdelitev verige $\{X_n\}$, ki je π-ireducibilna. Veriga $\{X_n\}$ je periodična s periodo $d \geq 2$, če obstajajo paroma disjunktne Borelove množice $E_0 \dots E_{d-1}$, za katere velja

$$\forall i \in \{0 \dots d-1\}\ \forall x \in E_i:\ P(E_{i+1 \bmod d} \mid x) = 1.$$

Pripomnimo, da tedaj sledi

$$\forall x \in E_i:\ P^d(E_i \mid x) = 1.$$

(Naj bo $x \in E_i$. Tedaj

$$\begin{aligned} P^2(E_{i+2 \bmod d} \mid x) &= \int_\Sigma P(E_{i+2 \bmod d} \mid y) dP(y \mid x) \\ &\geq \int_{y \in E_{i+1 \bmod d}} \dots \\ &= \int_{y \in E_{i+1 \bmod d}} dP(y \mid x) = 1.) \end{aligned}$$

Če veriga ni periodična, je aperiodična.

Trditev 4.2.3. Če sta $f(y)$ in $q(y \mid x)$ pozitivni in zvezni povsod na $\Sigma = \mathbb{R}^r$ oz $\Sigma \times \Sigma = \mathbb{R}^r \times \mathbb{R}^r$, je π-ireducibilna M-H veriga tudi aperiodična.

Izrek 4.2.4. Če je M-H veriga π-ireducibilna in aperiodična, velja

$$\lim_{n \to \infty} \sup_{A \in B(\mathbb{R}^r)} |P^n(A \mid x) - \pi(A)| = 0$$

$\pi = s.s.[x]$ (za s.g. vsak x).

Izrek 4.2.5. Privzemimo $q(y \mid x) = q(y)$ za $\forall x$ (neodvisni M-H) in $q(y) > 0$ ter $f(y) > 0$. Če obstaja konstanta M, za katero je $\forall y: \ f(y) < Mq(y)$, je M-H veriga ENAKOMERNO ERGODIČNA, obstaja zaporedje $r(n) \overset{n \to \infty}{\to} 0$, za katero velja $\sup_A |P^n(A \mid x) - \pi(A)| \leq r(n)$. Tu obstaja eksplicitni izraz za $r(n)$.

Opomba. Pri predpostavkah izreka zmano s A/R simulirati NEODVISNO vzorčenje iz π.

Definicija 4.2.6. Naj bo π stacionarna oprazdelitev verige $\{X_n\}$, ki je π-ireducibilna. Tedaj je ta veriga POVRNLJIVA, če iz $\pi(A) > 0$ sledi

(i) $P(X_n \in A$ neskončno mnogokrat $\mid x) > 0$ za $\forall x$,

(ii) $P(X_n \in A$ neskončno mnogokrat $\mid x) = 1$ za π-skoraj vse x.

Definicija 4.2.7. Veriga je Harrisov povrnljiva, če velja (ii) za $\forall x$.

Izrek 4.2.8. Naj bo π stacionarna porazdelitev verige $\{X_n\}$, ki je π-ireducibilna. Tedaj je veriga povrnljiva.

- Če je veriga še aperiodična, velja $\lim_{n\to\infty} \sup_A |P^n(A \mid x) - \pi(A)| = 0$ za vsak x (π-s.g. $[x]$).
- Če je veriga Harrisov povrnljiva, velja $\lim_{n\to\infty} \sup_A B\left(P^n(A \mid x) - \pi(A)\right) = 0$ za vsak x.

Poleg tega v tem primeru velja EZVŠ v obliki:
$\forall x$ $\forall$ integrabilna h:

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} h(x_i) = \int h(u)f(u)du\right) = P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} h(x_i) = \int h(u)d\pi(u) \mid x\right) = 1.$$

Trditev 4.2.9. π-ireducibilna M-H veriga je Harrisov povrnljiva.

Izrek 4.2.10. Privzemimo Metropolisovo verigo oblike $q(y \mid x) = q(y - x)$, kjer je q simetrična okrog 0. Naj bo $\Sigma = \mathbb{R}^r$. Tedaj pridružena veriga „nikdar" ni enakomerno ergodična.

4.3 Linearna regresija

4.3.1 Večrazsežna normalna porazdelitev

Pravimo $X \sim N(\mu, \Sigma)$ za $\mu \in \mathbb{R}^n$ in $\sigma \in P(n)$ (pozitivna definitna $n \times n$ matrika), če ima P_X Lebesgueovo gostoto

$$f(\mu, \Sigma) = (2\pi)^{-\frac{n}{2}}(det\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}\langle\Sigma^{-1}(x-\mu), x-\mu\rangle}.$$

Posebej $f(0, I) = (2\pi)^{-\frac{n}{2}}e^{-\frac{1}{2}\|x\|^2}$ je gostota t.i. STANDARDNE n-razsežne normalne porazdeliteve.
Transformacija. Če je $A : \mathbb{R}^n \to \mathbb{R}^m$ linearna surjekcija in je $X \sim N(\mu, \Sigma)$, je za $v \in \mathbb{R}^m$ $\quad Ax + v \sim N(A\mu + v, A\Sigma A^T)$ (tu vzamemo $A \in \mathbb{R}^{n\times n}$).
Posebej. Če je $Z \sim N(0, I)$ in je $\Sigma = AA^T$ (za $\Sigma, A \in \mathbb{R}^{n\times n}$), je $Az + \mu \sim N(\mu, \Sigma)$; torej vzorčenje $X \in N(\mu, \Sigma)$ realiziramo z vzorčenjem $z \in N(0, I_{n\times n})$ in transformacijo $x = Az + \mu$.

4.3.2 Klasična standardna linearna regresija in MNK

MNK: metoda najmanjših kvadratov.

Pri linearni regresiji skušamo „proučevano“ s.s. Y (ki jo je drago meriti ali pa je rezultat okoliščin v prihodnosti) napovedati na podlagi vrednosti „cenejših“ t.i. pojasnjevalnih / napovedovalnih s.s. $X_1 \dots X_p$ s pomočjo zveze

$$Y \approx \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p.$$

Zgled. Še vedno je v uporabi t.i. Frideualdova formula

$$LDL \approx THC - HDL - \frac{1}{2.4} TRI \quad \text{v mmd/l.}$$

Zgornjo linearno zvezo (v β) bi formalizirali v

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \epsilon,$$

kjer ima ϵ vlogo slučajnega odstopanja oz. meritvene napake.

Tega „modela“ ni mogoče parametrizirati zaradi (neznanih) porazdelitev prediktorja $X_1 \dots X_p$.

Da dobimo dober parametrični model, uvedemo fikcijo laboratorijskega eksperimenta

$$\begin{aligned}
Y_1 &= \beta_0 + X_{11}\beta_1 + \dots + X_{1p}\beta_p + \epsilon_1 \\
&\vdots \\
Y_n &= \beta_0 + X_{n1}\beta_1 + \dots + X_{np}\beta_p + \epsilon_n.
\end{aligned}$$

Tu do X_{ij} ($1 \leq i \leq n,\ 1 \leq j \leq p$) ZNANE (nastavljene) konstante, $\epsilon_1 \dots \epsilon_n$ pa so slučajne SPREMENLJIVKE.

Če označimo

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ & \vdots & & \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}_{n\times(p+1)} \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{p\times 1} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1}$$

sledi enačba linearne regresije $y = X\beta + \epsilon$.
<u>MNK</u>
Privzemimo meritev y; pri danem vektorju $\hat{\beta}$ tvorimo „napoved“ $\hat{y} = X\hat{\beta}$.
Kvaliteto vektorje $\hat{\beta}$ merimo preko

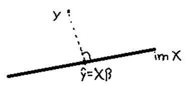
$$\|\hat{y} - y\|^2 = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 = VKR = VKR(\beta)$$

VKR: vsota kvadratov residualov.
Pišimo

$$VKR(\beta) = \|y - X\beta\|^2 = \langle y - X\beta, y - X\beta \rangle = \langle X\beta, X\beta \rangle - 2\langle y, X\beta \rangle + \langle y, y \rangle,$$
$$dVKR(\beta) = \beta^T X^t X - 2y^T X.$$

Pogoj za ekstrem je

$$X^T(y - X\beta) = 0 \text{ oz. } X^T X = X^T y.$$



Privzemimo $p + 1 \leq n$; tedaj je $X^T X$ obrnljva matrika natanko tedaj, ko je $rankX = p + 1$. Tedaj ima v $\hat{\beta} = (X^T X)^{-1} X^T y$ funkcija VKR (enoličen) minimum.
Pravimo, da je $\hat{\beta} = \hat{\beta}(y)$ ocena za β po MNK.

Zadnjič: ocena za β po MNK (linearna algebra).
Verjetnostni model(i) standardne linearne regresije:
$y = X\beta + \epsilon$;
y: proučevani slučajni vektor,
$X \in \mathbb{R}^{n \times d}$ matrika konstant,
$\beta \in \mathbb{R}^d$ (delni) parameter,
ϵ slučajen s porazdelitvijo v nekem modelu.
Vedno privzamemo $E(\epsilon) = 0 \in \mathbb{R}^m$ (če $\neq 0$: sistemska napaka).

Za variančno-kovariančno matriko $Var(\epsilon)$ tipično prizamemo, da je
$Var(\epsilon) = \sigma^2 I$ (homoskelastičnost + nekoreliranost).
Boljše modele dobimo z dodatnimi zahtevami na ϵ.
Tedaj postane $\hat{\beta} = \hat{\beta}(y) = (X^T X)^{-1} X^T y$ slučajni vektor, ki je cenilka za β.
(Privzamemo $rang X = d \leq n$; X polnega ranga, $(X^T X)^{-1}$ $\exists$).
Velja $E\left(\hat{\beta}(y)\right) = (X^T X)^{-1} X^T E(Y) = (X^T X)^{-1} X^T X \beta = \beta$; torej je β nepristranska (linearna) cenilka za β.

Zgled.

$X = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n, \beta \in \mathbb{R} (d = 1)$; pišimo $Var(\epsilon_i) = \sigma^2$
(vzamemo $Var(\epsilon) = \sigma^2 I$).
Tedaj je $Y_i = \mu + \epsilon_i,\ E(Y_i) = \mu, Var(Y_i) = \sigma^2$ ($\implies$ NEP vzorčenje).
V tem modelu (vemo) je standardna cenilka za σ^2 seveda
$S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \overline{y})^2$.
Očitno $\hat{\mu}(y) = \overline{y}$ (vstavimo β).

Standardna cenilka za σ^2 je $\sigma^2 = \frac{1}{n-d} \|y - X\hat{\beta}\|^2$ ($X\hat{\beta}$ „vzame“ d prostorskih stopenj).
Trdimo, da je nepristranska pri pogojih $E(\epsilon) = 0, Var(\epsilon) = \sigma^2 I$.
Za dokaz najprej razcepimo $X = S \cdot P$, kjer je $S \in \mathbb{R}^{n \times d}$ matrika z ortonormiranimi stolpci, $P \in \mathbb{R}^{d \times d}$ obrnljiva („neprehodna“) matrika.
To pomeni, da stolpci matrike S tvorijo ON bazo za $im\, X$.
Najpreprosteje je izpeljati GS ortogonalizacijo na stolpcih matrike X.
$S^1 = \frac{X^1}{\|X^1\|}$
$S^2 = \frac{X^2 - \langle X^2, S^1 \rangle S^1}{\|...\|}$.

Iz $\begin{bmatrix} X^1 & \dots & X^d \end{bmatrix} = \begin{bmatrix} S^1 & \dots & S^d \end{bmatrix} \begin{bmatrix} \|X^1\|^2 & * & \dots & * \\ & \|\dots\| & & * \\ & & \|\| > 0 & \\ \vdots & & & \\ & & & \|\| > 0 \end{bmatrix}$

preberemo, da je pri („direktni“) GS ortogonalizaciji P zgornje trikotna ma-

trika s pozitivnimi diagonalnimi elementi.
Izračunajmo $X^TX = P^TS^TSP = P^TI_{d\times d}P = P^TP$; sledi
$\hat{\beta} = (P^{-1}P^{-T}P^TS^Ty) = P^{-1}S^Ty$ in
$X\hat{\beta} = SS^Ty$ (pravokotna projekcija na X, SS^T projektor: $(SS^T)(SS^T) = SS^T$).
Zato je $y - X\hat{\beta} = (I - SS^T)y$; velja $E(y - X\hat{\beta}) = 0 \in \mathbb{R}^n$.
Pripomnimo, da za $\forall a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n$ velja $\|a^2\| = \sum_{a_i^2} = tr(aa^T)$.
Zato lahko izrazimo

$$\begin{aligned}
E(VKR) &= E(tr((y - X\hat{\beta})(y - X\hat{\beta})^T)) \\
&= tr(Var(y - X\hat{\beta})) \\
&= tr(Var((I - SS^T)y)) \\
&= tr((I - SS^T)Var(Y)(I - SS^T)^T) \\
&= \sigma^2 tr(I) + \sigma^2 tr(SS^T) \\
&= \sigma^2(n - d).
\end{aligned}$$

Normalni standardni linearni model
$Y = X\beta + \epsilon$, kjer je $\epsilon \in N(0, \sigma^2 I)$.
To je parametrični model s $\Theta = \mathbb{R}^d \times (0, \infty) = (\beta, \sigma^2)$.
Seveda sledi $Y \sim N(X\beta, \sigma^2 I)$.
Lastnosti cenilk $\hat{\beta}, \hat{\sigma^2}$:

- seveda je $\hat{\beta}$ normalno porazdeljen;
  velja $Var(\hat{\beta}) = (X^TX)^{-1}X^TVar(\beta)\left((X^TX)^{-1}X^T\right)^{-1} = \sigma^2(X^TX)^{-1}$
  $\implies \beta \sim N\left(\beta, \sigma^2(X^TX)^{-1}\right) = N(\beta, \sigma^2P^{-1}P^{-T})$,

- dopolnimo S do ortogonalne matrike $Q = \begin{bmatrix} S' & S \end{bmatrix} \in O^{n\times n}$ (ortogonalna $n \times n$ matrika) (tu je S matrika $\in \mathbb{R}^{n\times(n-d)}$ z ON stoplci, ki razpenjajo

$(im\,, X)^{\perp})$

$$\begin{aligned} VKR &= \|y - X\hat{\beta}\|^2 \\ &= \langle (I - SS^T)y, (I - SS^T)y \rangle \\ &\overset{\text{projekcija}}{=} \langle (I - SS^T)y, y \rangle \\ &\overset{\text{ortog.}}{=} \langle Q^T(I - SS^T)(QQ^T)y, Q^T y \rangle \end{aligned}$$

$$\text{(i)}\ \ Q^T S = \begin{bmatrix} (S')^T \\ S^T \end{bmatrix} S = \begin{bmatrix} (S')^T S \\ S^T S \end{bmatrix} = \begin{bmatrix} 0 \\ I \end{bmatrix}$$

$$(Q^T S)(Q^T S)^T = \begin{bmatrix} 0 \\ I \end{bmatrix} \begin{bmatrix} 0 & I \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & I_{d\times d} \end{bmatrix}$$

$$\implies Q^T(I - SS^T)Q = \begin{bmatrix} I_{(n-d)\times(n-d)} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\text{(ii)}\ \ Q^T y \sim N\left(\begin{bmatrix} (S')^T \\ S^T \end{bmatrix} X\beta, Q^T \sigma^2 I Q \right) = N\left(\begin{bmatrix} 0_{n-d} \\ P\beta \end{bmatrix}, \sigma^2 I \right).$$

Pišimo $z = Q^T y = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$,

sledi $\frac{VKR}{\sigma^2} = \sum_{i=1}^{n-d} \frac{z_i^2}{\sigma^2}$ in zato je

$\frac{VKR}{\sigma^2} \sim \chi^2_{n-d} = \text{Gama}\left(\frac{n-d}{2}, \frac{1}{2}\right)$.

4.3.3 Standardna neinformativna normalna porazdelitev

(Uporabljamo pri t.i. večkratni imputaciji.)

Gre za model $(X \sim \beta, \sigma^2) \sim N(X\beta, \sigma^2 I)$ z NEPRAVO apriorno gostoto $f(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$ (to je „ploščata“ gostota v spremenljivki $\ln(\sigma^2)$ (vaja/dn)).

Vzorčna gostota je

$$f(y \mid \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\|y - X\beta\|^2}.$$

Razcepimo

$\|y - X\beta\|^2 = \|y - -X\hat{\beta} + X\hat{\beta} + X\beta\|^2 = \|y - X\hat{\beta}\|^2 + \langle X(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle;$

$X\hat{\beta} - X\beta \in im\, X,\ y - X\hat{\beta} \in (im\, X^{\perp})$.
Aposteriorna gostota

$$\begin{aligned} f(\beta, \sigma^2 \mid y) &\propto (\sigma^2)^{-1-\frac{n}{2}} e^{-\frac{1}{\sigma^2}\cdot\frac{VKR}{2} - \frac{1}{2}\langle \frac{X^TX}{\sigma^2}(\hat{\beta}-\beta), \hat{\beta}-\beta \rangle} \\ &\propto (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{VKR}{2\sigma^2}} (\sigma^2)^{\frac{d}{2}} \sqrt{det\left(\frac{X^TX}{\sigma^2}\right)} e^{-\frac{1}{2}\langle \frac{X^TX}{\sigma^2}(\beta-\hat{\beta}), \beta-\hat{(\beta)}\rangle}. \end{aligned}$$

$\frac{X^TX}{\sigma^2} = (\text{var-kovar})^{-1}$.
Vidimo: $\int_{\sigma^2}\int_{\beta\in\mathbb{R}^d}\cdots < \infty$, če $n-d>0$.
Sledi, da imamo pri $n-d>0$ prabo aposteriorno porazdelitev

$$f(\beta, \sigma^2 \mid y) = f_{\text{InvGama}\left(\frac{n-d}{2}, \frac{VKR}{2}\right)}(\sigma^2) \cdot f_{N(\hat{\beta}, \sigma^2(X^TX)^{-1})}(\beta).$$

V praksi mas zanima VZORČENJE iz te aposteriorne porazdelitve.
Najprej vzorčimo $\sigma^2 \in \text{InvGama}\left(\frac{n-d}{2}, \frac{VKR(y)}{2}\right)$, potem vzorčimo (pogojno na dobljeni σ^2) iz
$N(\hat{\beta}, \sigma^2(X^TX)^{-1}) = N(\hat{\beta}, \sigma^2 P^{-1}P^{-T})$.
To realiziramo na slednji način:

(i) vzorčimo $g \in \chi^2_{n-d}$, za realizacijo $\sigma^2 \in \text{InvGama}\left(\frac{n-d}{2}, \frac{VKR}{2}\right)$ vzamemo $\frac{VKR}{g}$,

(ii) vzorčimo $z \in N(0, I_{d\times d})$ in za realizacijo
$\beta \in (\beta, \sigma^2, y)$ vzamemo

$$\begin{aligned} \hat{\beta} + \sigma^2 P^{-1} z &= P^{-1}S^Ty + \sqrt{\frac{VKR(y)}{g}} P^{-1}z \\ &= P^{-1}\left(S^Ty + \sqrt{\frac{VKR(y)}{g}} z\right). \end{aligned}$$

Ta postopek je numerično učinkovit.
Pripomnimo še: $E(\text{InvGama}(a,b)) = \frac{b}{a-1}$
$\implies E\left(\text{InvGama}\left(\frac{n-d}{2}, \frac{VKR}{2}\right)\right) = \frac{\frac{VKR}{2}}{\frac{n-d}{2}-1} = \frac{VKR}{n-d-2} \overset{\text{asimptotično}}{\sim} \frac{VKR}{n-d}$ - nepristranska.

4.3.4 Bayesovi modeli linearne regresije z znano varianco

(in informativno apriorno porazdelitvijo v β).

$(y \mid \beta) \sim N(X\beta, \sigma^2 I)$

Izkaže se, da je konjugirana družina porazdelitev ravno
$\{N(\beta_0, T_0) \mid \beta_0 \in \mathbb{R}^d, T_0 \in P(d)\}$.
V praksi pogosto uoprabljamo tudi konjugirano (pod)družino
$\{N\left(\beta_0, \tau_0^2 (X^T X)^{-1}\right) \mid \beta_0 \in \mathbb{R}^d, \tau_0 \in (0, \infty)\}$.
Aposteriorna porazdelitev
$N(\beta_1, T_1)$, kjer je

$$\begin{aligned} T_1^{-1} &= T_0^{-1} + \frac{X^T X}{\sigma^2} \\ &= \frac{X^T X}{\tau_0^2} + \frac{X^T X}{\sigma^2} \\ &= \left(\frac{1}{\sigma^2} + \frac{1}{\tau_0^2}\right) X^T X \end{aligned}$$

frekventistična preciznost vzorca in

$$\begin{aligned} \beta_1 &= T_1 (T_0^{-1} \beta_0 + (X^T X)^{-1} (X^T X) X^T y) \\ &= \left(T_0^{-1} + \frac{X^T X}{\sigma^2}\right)^{-1} \cdot \left(T_0^{-1} \beta_0 + \frac{X^T X}{\sigma^2} \hat{\beta}\right) \\ &= \left(T_0^{-1} + \frac{X^T X}{\sigma^2}\right)^{-1} \cdot \left(\frac{1}{\tau_0^2} \beta_0 + \frac{1}{\sigma^2} \hat{\beta}\right) \end{aligned}$$

(daljica med β_0 in $\hat{\beta}$)
($\equiv \frac{a\beta_0 + b\hat{\beta}}{a+b}$.)

$(Y \mid \beta) \sim N(X\beta, V)$

V je znana vzorčno-kovariančna matrika.
Za $Z = V^{-\frac{1}{2}} Y$ velja $(Z \mid \beta) \sim N(V^{-\frac{1}{2}} X\beta, I)$;
$W = V^{-\frac{1}{2}} X\beta$.

Ocena za β po MNK je zato
$\widehat{\beta(z)} = (W^T W)^{-1} W^T z$ oz.
$\widehat{\beta(y)} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$ (UNK - uteženi najmanjši kvadrati).
Konjugirana družina apriornih porazdelitev je seveda normalna
$\{N(\beta_0, T_0) \mid \beta_0 \in \mathbb{R}^d, T_0 \in P(d)\}$.
Aposteriorna (apriorna?) porazdelitev pri opaženem y je tedaj $N(\beta_1, T_1)$, kjer
$T_1^{-1} = T_0^{-1} + X^T V^{-1} x$ in
$\beta_1 = (T_0^{-1} + X^T V^{-1} x)(T_0^{-1} \beta_0 + X^T V^{-1} y)$.

4.3.5 Vključevanje apriorne informacije glede variance

Uvodoma si oglejmo model $(Y \mid \sigma^2) \sim N(X\beta, \sigma^2 I)$

Velja $f(y \mid \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\|y - X\beta\|^2}$.
Konjugirana družina apriornih porazdelitev je očitno InvGama.
Pri $f(\sigma^2) = f_{\mathrm{InvGama}(a,b)}(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \cdot e^{-\frac{b}{\sigma^2}}$
realizacije y je aposteriorna porazdelitev
$\mathrm{InvGama}(a + \frac{n}{2}, b + \frac{1}{2}\|y - X\beta\|^2)$.

$(Y \mid \beta, \sigma^2) \sim N(X\beta, \sigma^2 I)$ **s polkonjugirano apriorno porazdelitvijo**

Tu je $f(\beta \mid \sigma^2) = f_{\mathrm{konj.}}(\beta) \cdot f_{\mathrm{konj.}}(\sigma^2) = f_{N(\beta_0, T_0)}(\beta) \cdot f_{\mathrm{InvGama}(a,b)}(\sigma^2)$.
Uporabimo *Bayesovi modeli linearne regresije z znano varianco* in *Uvodoma si oglejmo model ..* in sledi
$f(\beta \mid \sigma^2, y) = f_{N(\beta_1, T_1)}(\beta)$ in
$f(\sigma^2 \mid \beta, y) = f_{\mathrm{InvGama}(a + \frac{n}{2}, b + \frac{1}{2}\|y - X\beta\|^2)}(\sigma^2)$
za
$T_1^{-1} = T_0^{-1} + \frac{X^T X}{\sigma^2}$ in
$\beta_1 = (T_0^{-1} + \frac{X^T X}{\sigma^2})^{-1} (T_0 \beta_0 + \frac{X^T Y}{\sigma^2})$.
Do polne aposteriorne porazdelitve $f(\beta, \sigma^2 \mid y)$ dostopamo preko MCMC z Gibbsovim vzorčevalnikom.

$(Y \mid \beta, \sigma^2) \sim N(X\beta, \sigma^2 I)$ s konjugirano apriorno porazdelitvijo

Izkaže se, da je konjugirana družina podana z gostoto
$f(\beta, \sigma^2) = f(\beta \mid \sigma^2) \cdot f(\sigma^2) = f_{N(\beta_0, \sigma^2\Lambda_0)}(\beta) \cdot f_{\mathrm{InvGama}(a,b)}(\sigma^2)$.
(Parametri konjugirane družine so $\beta_0 \in \mathbb{R}^d, \Lambda_0 \in P(d); a, b \in (0, \infty)$.)
Tudi aposteriorno porazdelitev pri y predstavimo dvofazno
$f(\beta, \sigma^2 \mid y) = f(\beta \mid \sigma^2, y) \cdot f(\sigma^2 \mid y)$.
Velja $f(\beta \mid \sigma^2, y) = f_{N(\beta_1, \sigma^2\Lambda_1)}(\beta)$, kjer
$\Lambda_1^{-1} = \Lambda_0^{-1} + X^T X$ in
$\beta_1 = (\Lambda_0^{-1} + X^T X)^{-1}(\Lambda_0^{-1}\beta_0 + X^T y)$
in
$f(\sigma^2 \mid \beta, y) = f_{\mathrm{InvGama}(a_1, b_1)}(\sigma^2)$, kjer
$a_1 = a + \frac{n}{2}$ in
$b_1 = b + \frac{1}{2}\langle (I + X\Lambda_0 X^T)^{-1}(y - X\beta_0), y - X\beta_0 \rangle$
$= b + \frac{1}{2}\left(\langle y, y\rangle + \langle \Lambda_0\beta_0, \beta_0\rangle - \langle \Lambda_1\beta_1, \beta_1\rangle\right)$.