

# Bayesova statistika - zapiski s predavanj prof. Smrekarja

Tomaž Poljanšek

študijsko leto 2023/24

# Kazalo

<b>1</b>	<b>Uvod</b>	<b>1</b>
1.1	Elementarna Bayesova statistika . . . . .	1
1.2	Proučevani slučajni vektor (vzročni) parametrični model . . .	2
1.3	Apriorna in „robna“ porazdelitev . . . . .	3
1.4	Disperzija aposteriornih porazdelitev . . . . .	6
<b>2</b>	<b>Enoparametrični modeli</b>	<b>10</b>
2.1	Beta-binomski model . . . . .	10
2.2	Poissonov model (gama-poissonov model) . . . . .	10
2.3	Normalni model z znano disperzijo . . . . .	11
2.4	Ekspontentne družine porazdelitev . . . . .	13
2.5	Neinformirane apriorne porazdelitve . . . . .	15
<b>3</b>	<b>Monte-Carlo integracija in metode vzorčenja</b>	<b>18</b>
3.1	Klasična integracija Monte-Carlo . . . . .	19
3.2	Simulacija vzorčenja z inverzno kumulativno porazdelitveno funkcijo . . . . .	20
3.3	Metoda sprejmi ali zavrni (A/R) . . . . .	21
3.4	Metode MCMC (Monte Carlo z markovskimi verigami) . . . .	23

## Seznam uporabljenih kratic

kratica	pomen
<b>s.v.</b>	slučajni vektor
<b>B</b>	binomska porazdelitev
<b>NEP</b>	neodvisen in enako porazdeljen
<b>s.s.</b>	slučajna spremenljivka
<b>p.v.</b>	pričakovana vrednost
<b>AKI</b>	aposteriorni kredibilnostni interval
<b>BF</b>	Bayesova formula
<b>s.g.</b>	skoraj gotovo
<b>k.p.f.</b>	kumulativna porazdelitvena funkcija
<b>A/R</b>	accept or reject
<b>M.v.</b>	Markovska veriga
<b>EZVŠ</b>	ergodični zakon velikih števil
<b>ECLI</b>	ergodični CLI

# 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  $E_1, E_2 \dots E_m : E_i \cap E_j = \emptyset$  za  $i \neq j$  in  $E_1 \cup E_2 \cup \dots \cup E_m = \Omega$ .

Če imamo še neki dogodek  $A$ , velja t.i. zakon o popolni verjetnosti

$P(A) = \sum_{i=1}^m P(A | E_i) \cdot P(E_i)$  (interpretacija: 2-fazni poskus).

V Bayesovem okviru nas zanimajo  $P(E_j | A)$  (verjetnost, da se je v „1. fazi“ zgodil  $E_j$ , če se je „2. fazi“ zgodil  $A$ ). Ker je

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

je

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

oziroma

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

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

Informacijo  $\theta \approx 400$  zakodiramo kot  $E(\Theta) = 400$ .

$$\begin{aligned} \text{Privzamemo (kar!) } \Theta &\sim B\left(1000, \frac{4}{10}\right) \\ \Rightarrow P(\Theta = \theta) &= \binom{1000}{\theta} \left(\frac{4}{10}\right)^\theta \left(1 - \frac{4}{10}\right)^{1000-\theta} \\ P(k \text{ črnih od } 10 \text{ izvlečenih} | \Theta = \theta) &= \frac{\binom{\theta}{k} \binom{1000-\theta}{10-k}}{\binom{1000}{k}} \quad (*) \end{aligned}$$

(\*) pri omejitvah ( $k$  omejimo).

Osvežena porazdelitev - novo vedenje

$$\begin{aligned} P(\Theta = \theta | 6 \text{ črnih od } 10 \text{ izvlečenih}) &= \\ \frac{P(6 \text{ črnih od } 10 \text{ izvlečenih} | \Theta = \theta) \cdot P(B(1000, \frac{4}{10}) = \theta)}{\sum_{i=0}^{1000} P(6 \text{ črnih od } 10 \text{ izvlečenih} | \Theta = i) \cdot P(B(1000, \frac{4}{10}) = i)} &= i \end{aligned}$$

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  $X$ . Zanj 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.)  $\Theta$  z vrednostmi v  $\Theta$  (večinoma  $r \geq 2$ ).

Porazelitvi s.v.  $X_i$  pogojno na  $\Theta = \theta$  pravimo vzorčna porazdelitev. Privzeli bomo, da imamo gostote  $f(x | \theta)$  ali verjetnostne funkcije

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

torej da velja

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

(v Lebesgueovi meri) ali

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

Modelu pogojnih porazdelitev  $(X | \Theta = \theta)$  pravimo vzorčni model.

### 1.3 Apriorna in „robna“ porazdelitev

Porazdelitvi fiktivnega slučajnega vektorja  $\Theta$  pravimo apriorna porazdelitev, brezpogojni (robni) porazdelitveni slučajnega vektorja  $X$  pa pravimo „robna“ porazdelitev

(\*) v resnici sta obe porazdelitvi robni porazdelitvi družne porazdelitve vektorja  $(X, \Theta)$  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.)  $\Pi$  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

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

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 ( $P(\text{cifra} = p)$ ), gre za slučajne spremenljivke  $X_1, X_2 \dots X_n$ , za katere je  $(X_i \mid \Pi = p) \sim \text{Bernoulli}(p)$  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

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

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

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

za  $i \neq j$ , torej so  $X_i$  brezpogojno neodvisne  $\iff \Pi = \text{konstantna}$  (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  $f(p) = f_{Beta(a,b)}(p)$ )

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

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

Aposteriorna pričakovana vrednost (p.v.) je

$$\begin{aligned} \frac{a+k}{a+b+n} &= \frac{(a+b) \frac{a}{a+b} + n \frac{k}{n}}{a+b+n} = \\ &= \frac{a+b}{a+b+n} \cdot \frac{a}{a+b} + \frac{n}{a+b+n} \cdot \frac{k}{n}. \end{aligned}$$

Tukaj je



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

Vzorec velik  $\rightarrow$  prevlada mnenje vzorca.

## 1.4 Disperzija aposteriornih porazdelitev

Gre pravzaprav za disperzijo pogojnih porazdelitev. Naj bosta  $X : \Omega \rightarrow \mathbb{R}^m$  in  $Y : \Omega \rightarrow \mathbb{R}^n$  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  $\mu$  in  $f_Y(y) = \int f_{(X,Y)}(x, y) d\mu(x)$  za  $Y$  glede na  $\nu$ . 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  $\nu$ : gostota v  $X \rightarrow \mu$  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.**

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

Slučajni vektor  $\rightarrow$  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.

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

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

$$\text{Var}(Y \mid X) = v(X).$$

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“

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

kjer sta seveda

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

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

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

$$E(\text{Var}(\Pi \mid T)) \leq \text{Var}(\Pi).$$

*Opomba.*

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

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

(V povprečju aposteriorna varianca boljša.)

### 1.4.1 Aposteriorni kredibilnostni interval

(Bayesova različica intervala zaupanja).

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

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

$p$  je slučajen.

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

$$\begin{aligned} L(k) &= F_{Beta(a+k, b+n-k)}^{-1}\left(\frac{\alpha}{2}\right), \\ U(k) &= F_{Beta(a+k, b+n-k)}^{-1}\left(1 - \frac{\alpha}{2}\right) \end{aligned}$$

in v aposteriori 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.

### 1.4.2 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.

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

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

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

$f(x)$  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  $(X_i \mid \lambda) \stackrel{\text{NEP}}{\sim} \text{Poisson}(\lambda)$ . Za proučevano s.s. vzamemo  $T = \sum_{i=1}^n X_i$ ; seveda je  $(T \mid \lambda) \sim \text{Poisson}(n\lambda)$ .

Privzemimo apriorno porazdelitev

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

$(0,\infty)$  je parametrični prostor,  $a$  in  $b$  sta pozitivni konstanti.

Izkaže se:

$$\begin{aligned} E(\text{Gama}(a,b)) &= \frac{a}{b}, \\ D(\text{Gama}(a,b)) &= \frac{a}{b^2}. \end{aligned}$$

DN.

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

Bayesova formula se glasi

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

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

**Definicija 2.2.1.** Naj bo podan vzorčni model  $P$  in naj bo  $K$  družina porazdelitev na parametričnem prostoru  $\Theta$ . 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  $\Theta$ . Rečemo lahko tudi, da sta  $P$  in  $K$  konjugiran par.

## 2.3 Normalni model z znano disperzijo

Tu je  $\sigma^2$  znana disperzija, vzorec  $X = (X_1 \dots X_n)$  pa zadošča  $(X_i \mid \mu) \stackrel{\text{NEP}}{\sim} N(\mu, \sigma^2)$ , 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

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

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

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

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

Velja

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

Sledi

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

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 porazdelitev seštevajo.

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

To pomeni

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

in

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

## 2.4 Eksponentne družine porazdelitev

Vzorčni model pripada eksponentni družini porazdelitev, če velja

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

kjer je

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

*Zgled.*

① NEP Bernoullijev model:

$$\begin{aligned}
 (X_i \mid p) &\stackrel{\text{NEP}}{\sim} \text{Bernoulli}(p) = B(1, p) \\
 f(x_1 \dots x_n \mid p) &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \cdot 1_{\{0,1\}^n}(x_1 \dots x_n) \\
 f(x_1 \dots x_n \mid p) &= P(X_1 = x_1 \dots X_n = x_n \mid p)
 \end{aligned}$$



Preoblikujemo v:

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

② Normalni model z znano  $\sigma^2$ :

$$\Theta = \mathbb{R}, \mu \in \mathbb{R}$$

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

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

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

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

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

Tukaj je  $\mu \in \mathbb{R}^n$  in  $v \in \mathbb{R}$ .

(Upoštevati moramo morebitne omejitve zaradi zahteve  $\int F = 1$ ).

Seveda je

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

Aposteriorna gostota?

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

Tukaj smo zmožili nekonstantne faktorje iz 2.1 in 2.2.

Vidimo:

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

Gre za konjugirano družino.

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

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

kjer sta  $\eta, v \in \mathbb{R}$ .

Vidimo, da mora biti  $v > 0$ .

DN:  $\eta, v \rightarrow \mu_0, \tau_0^2$  - 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  $U(0,1) = \text{Beta}(1,1)$  kot neinformirano porazdelitev.

Učinek reparametrizacije

*Zgled.* Binomski vzorčni model:  $f(k | p) = \binom{n}{k} p^k (1-p)^{n-k}$ ,  $p \in (0,1)$ .

Reparametrizirajmo s parametrom  $q = \ln\left(\frac{p}{1-p}\right) = \text{logit}(p)$ . Dobimo

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

Kako je s transformacijo apriorne porazdelitve?

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

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

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

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

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

$$\begin{aligned} FI(\theta) &= E_{(X|\Theta)}(\text{grad}_{\theta} \ln(f(x | \theta)) \cdot \text{grad}_{\theta} \ln(f(x | \theta))^T) \\ &= \text{Var}_{(X|\Theta)}(\text{grad}_{\theta} \ln(f(x | \theta))) \\ &= -E_{(X|\Theta)}(H_{\theta}(\ln f)(x | \theta)). \end{aligned}$$

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

Učinek reparametrizacije na Jeffreysovo apriorno porazdelitev

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

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

diferenciabilen; sledi

$$\tilde{f}(x | \lambda) = f(x | \phi^{-1}(\lambda)).$$

Odvajajmo

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

$$\text{grad}_{\lambda} f(x | \lambda) = [J(\phi^{-1}(\lambda))]^T \cdot \text{grad}_{\theta} \ln f(x | \phi^{-1}(\lambda))$$

$J$ : Jacobijeva matrika.

$$\begin{aligned} \ln \tilde{f}(x | \lambda) &= \ln(f(x | \phi^{-1}(\lambda))) \cdot \text{grad}_{\lambda} \ln(f(x | \lambda)) \cdot \text{grad}_{\lambda} \ln(f(x | \lambda))^T \\ &= [J\phi^{-1}(\lambda)]^T \cdot \text{grad}_{\theta} \ln f(x | \phi^{-1}(\lambda)) \cdot \text{grad}_{\theta} \ln f(x | \phi^{-1}(\lambda))^T \cdot [J\phi^{-1}(\lambda)] \\ &\implies \widetilde{FI}(\lambda) = (J\phi^{-1}(\lambda))^T \cdot FI(\phi^{-1}(\lambda)) \cdot J\phi^{-1}(\lambda). \end{aligned}$$

Kaj je Jeffreysova apriorna porazdelitev na  $\lambda$ ?

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

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

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

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

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

Problemi:

- $h(\theta)f(\theta | x)$  je lahko zahtevna za integracijo - numerično,
- morda je „že“  $f(x | \theta)f(\theta)$  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  $\mu$ , velja

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

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

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

skoraj gotovo.

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

$$\frac{h(x_1) + \dots + h(x_n)}{n} \rightarrow \int h(x) \cdot | \cdot | dx$$

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  $\cdot | \cdot$ .

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

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

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.)

$F : \mathbb{R} \rightarrow [0,1]$  zvezna bijekcija  $\mathbb{R} \rightarrow (0,1)$ . 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“  $\Phi^{-1}$ , 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  $N(0,1)$  (  $\xrightarrow{\text{vaja}}$  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)  $F^{-1}(U) \sim F$ . (  $\xrightarrow{\text{vaja}}$  znamo vzorčiti iz končnih diskretnih porazdelitev).

### 3.3 Metoda sprejmi ali zavrni (A/R)

Motivacija: Naj bo  $f : \mathbb{R} \rightarrow [0, \infty)$  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

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

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  $g : \mathbb{R} \rightarrow [0, \infty)$  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

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

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  $g = F_{Cauchy}$  ( $g(x) = \frac{1}{\pi(1+x^2)}$ ).

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

$u$  smo izrazili iz  $x = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$ .

Vzorčenje iz Cauchyja je  $\tan\left(\pi\left(U - \frac{1}{\pi}\right)\right)$ ,  $U$  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

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

dobro funkcionira tudi v primeru, ko je  $X^{(0)}, X^{(1)}, X^{(2)} \dots$  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)  $\forall n :$

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

- 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

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

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  $\gamma_n$  (MCMC disperzija), za katero

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

(!!  $\gamma_n$  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  $\frac{\text{število členov (do } n\text{-tega, ki padejo v } A)}{n} \xrightarrow{n \rightarrow \infty} \int 1_A(x)f(x)dx = a$ .

### 3.4.1 Metropolisov algoritem

Naj bo  $f$  ciljna gostota in naj bo  $\{q(y | x) \mid y, x \text{ iz prostora stanj}\}$  družina „predlaganih“ gostot: za vsak  $x$  (iz katerih pa znamo simulirati NEP vzorčenje) je  $q(\_ | x)$  gostota neke porazdelitve. Naj bo še  $q(y | x) = q(x | y)$  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  $q(\_ | x^{(n-1)})$ .

Če velja  $f(y) \geq f(x^{(n-1)})$ , vzamemo  $X^{(n)} = y$  (realizacija).

Če je  $f(y) < f(x^{(n-1)})$ , vzamemo

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

Naenkrat.  $X^{(n)} = y$  z verjetnostjo  $\rho = \min\{\frac{f(y)}{f(x^{(n-1)})}, 1\}$  (\*)

(\*): če  $f(x^{(n-1)}) = 0$ , 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  $f(y) > 0 \implies \forall x : q(y | x) > 0$ , ima veriga EZVŠ.

Bayesova aplikacija.

Ciljna gostota je  $f(\theta | x)(v \theta)$ .

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

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

v resnici ne potrebujemo normalizacijske konstante v Bayesovi formuli.

Tipični primeri predlaganih gostot:

1.  $(Y | X^{(n-1)} = x^{(n-1)}) \sim U(K_\delta(x^{(n-1)}))$  za fiksni  $\delta$  (krogla s polmerom  $\delta$ ), v neki metiki.

(Povrnljivost - povsod, kjer neničelne verjetnosti, jemlje  $\infty$ -krat.)

(To pomeni  $q(y | x) = \frac{1}{Vol(K_\delta(x))} \cdot 1_{K_\delta(x)}(y)$  (namesto klasične uporabimo  $\infty$  metriko).)

2.  $(Y | X^{(n-1)} = x^{(n-1)}) \sim N(x^{(n-1)}, \Sigma)$  za fiksno  $\Sigma$ .

(To pomeni  $q(y | x) = (2\pi)^{-\frac{dim}{2}} (det \Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2} \langle \Sigma^{-1}(y-x), (y-x) \rangle}$  - simetričnost  $\checkmark$ .)

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  $\rho$  iz 3.4.1. uporabimo

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

(in  $\rho = 1$  če  $f(x^{(n-1)}) \cdot q(y \mid x^{(n-1)}) = 0$ ).

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  $\forall f(y) > 0$ .