

O'ZBEKISTON OLIY VA O'RTA MAXSUS TA'LIM VAZIRLIGI

MIRZO ULUG'BEK nomidagi O'ZBEKISTON MILLIY UNIVERSITETI

MEXANIKA-MATEMATIKA FAKULTETI

“EHTIMOLLAR NAZARIYASI VA MATEMATIK STATISTIKA”

KAFEDRASI

“Statistik baholashning EM algoritmi”

mavzusidagi

BITIRUV MALAKAVIY ISHI

Bajardi: “Matematika” ta'lim yo'nalishi

bitiruvchisi Ismatullayev N.N.

ILMIY RAHBAR:

prof. Abdushukurov A.A.

Bitiruv malakaviy ishi kafedradan dastlabki himoyadan o'tdi.

19 sonli bayonnoma “15” may 2014 y.

Toshkent – 2014

Mundarija

Kirish	3
I.Klassik modelda noma'lum parametrlarni baholash	5
1.1-§.Nuqtaviy baholash usullari va ularning hossalari.....	5
1.2-§.Muhim taqsimotlar noma'lum parametrlarining baholari va ularning hossalari.....	10
1.3-§.Nyuton-Rafson usuli.....	16
II.Noklassik modelda noma'lum parametrlarni baholash	21
2.1-§.To'liq bo'lmagan tanlanmalar modellari.....	21
2.2-§.Statistik baholashning EM algoritmi.....	22
2.3-§.I tur senzurlanish modelida gamma taqsimot noma'lum parametrlarni baholash.....	25
Xulosa	32
Foydalanilgan adabiyotlar	33

Kirish

Mazkur bitiruv malakaviy ishi EM algoritmini o`rganish va uni haqiqatga maksimal o`xshashlik bahosiga yaqinlashishini tekshirishga bag`ishlangan. Bitiruv malakaviy ishi kirish, 2 bob, 6 ta paragraf, xulosa va foydalanilgan adabiyotlar ro`yxatidan iborat. 1-bobda baho va uning xossalari, ba`zi muhim taqsimot funksiyalar orasidagi bog`lanishlar va noma`lum parametrlari baholari, hamda Nyuton-Rafson usuli yoritilgan. Bitiruv malakaviy ishi 2- bobida senzurlagan tanlanmalarning ba`zi modellari, EM algoritmi chuqur o`rganilgan, ya`ni EM algoritmi yordamida gamma taqsimotining noma`lum parametrlari uchun baho topilgan.

Bizga ma`lumki statistik tekshiruvlar davomida kuzatuvchi oldida bir qancha muammolar paydo bo`ladi. Yo`qotishga ega tizim yoki senzuralanish mavjudligida statistik bahoni olish juda qiyin, shuning uchun bunday hollarda iterativ usullardan foydalaniladi. Bunday iterativ usullardan biri Nyuton-Rafson usulining bir ko`rinishi bo`lgan EM algoritmidir. EM algoritmi inglizcha “Expectation maximization” “so`zlarining qisqartmasi bo`lib, matematik kutilmani maksimallashtirish algoritmi degan ma`noni anglatadi. EM algoritmi to`liqsiz ma`lumotlar masalalarida Xaqiqatga maksimal o`xshashlik bahosi uchun juda umumiy algoritmdir. EM algoritmi yordamida yechilishi mumkin bo`lgan masalalar ko`lami juda keng. EM algoritmida to`liqsiz ma`lumotlarni qayta ishlash g`oyasi formallashtirilgan bo`lib, u quyidagilardan iborat:

1. Yo`qotilgan ma`lumotlarni uning bahosi bilan to`ldirish.
2. Parametrlarni baholash.
3. Tushirib qoldirilgan qiymatlarni qayta baholash, bunda parametr baholari aniq deb hisoblanadi.

4. Parametrlar qayta baholanadi, jarayon yaqinlashishigacha davom ettiriladi.

To`liq ma`lumotlar uchun haqiqatga o`xshashlik funksiyasi logarifmi bo`yicha chiziqli bo`lgan jarayonlarni EM algoritmi hisoblaydi yoki umumiy holda ayrim yo`qotishlarni baholamasdan yo`qotilgan yetarli statistikani baholash.

Bundan tashqari algoritmning har iteratsiyasida $I(\theta|Y)$ (bu yerda Y -to'liq ma'lumotlar modeli, θ -parametr) haqiqatga o'xshashlik funksiyasi logarifmini baholash kerak. EM algoritmi yo'qotishni to'ldirish va iteratsiyani yakunlash go'yasi daslab hususiy hollarda taklif qilingan edi.

EM algoritmining har bir iteratsiyasida E (matematik kutilmani hisoblash) va M(maksimallashtirish) qadamlaridan tashkil topgan. Bu algoritmning muhim hususiyatlaridan biri unung ishonchli yaqinlashishi ya'ni aniq qat'iy bo'lmagan shartlar ostida har bir iteratsiya $I(\theta|Y_{kuz})$ bu yerda Y_{kuz} Y ning kuzatilayotgan qismi xaqiqatga o'xshashlik funksiyasi logarifmini kattalashtiradi. Agar $I(\theta|Y_{kuz})$ chegaralangan bo'lsa u holda $I(\theta^{(t)}|Y_{kuz})$ ketma-ketlik $I(\theta|Y_{kuz})$ ning statsionar qiymatiga yaqinlashadi. Umumiy holda aytish mumkinki $\theta^{(t)}$ ketma-ketlik yaqinlashuvchi bo'lsa, u holda ketma-ketlik lokal maksimumga yoki $I(\theta|Y_{kuz})$ egrilik nuqtasiga yaqinlashadi.

I-Bob. Klassik modelda nomalum parametrlarni baholash.

1.1-§ Nuqtaviy baholash usullari va ularning hossalari.

Statistik model $\{P_\theta, \theta \in \Theta\}$, $\Theta \subseteq R^{(s)}$, oila bilan berilgan bo'lib, $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ parametrlarni baholash masalasini qaraymiz.

Momentlar usuli. $G(\theta) = (g_1(\theta), \dots, g_s(\theta))$ vektor funksiya uchun biror asosli $\mathcal{G}_n^0(X^{(n)}) = (\mathcal{g}_{1n}^0(X^{(n)}), \dots, \mathcal{g}_{sn}^0(X^{(n)}))$ baho mavjud bo'lsin. Momentlar usuliga asosan, $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ uchun $\theta_n^0 = (\mathcal{g}_{1n}^0, \dots, \mathcal{g}_{sn}^0)$ baho sifatida $G(\theta) = \mathcal{G}_n^0$ tenglamaning, ya'ni

$$g_i(\theta) = \mathcal{g}_{in}^0(X^{(n)}), \quad i = \overline{1, s}; \quad (1.1.1)$$

sistemaning yechimi olinadi. Bunday baholarning xossalari $g_i, i = \overline{1, s}$, funksiyalarning xossalari bilan aniqlanadi. Odatda $g_i(\theta) = M_\theta a_i(\xi), i = \overline{1, s}$, (masalan $a(\xi) = \xi^i$) ko'rinishda tanlanadi. Bu holda katta sonlar qonunidan foydalanib, $\mathcal{g}_{in}^0$ sifatida $a_i(\xi)$ ning empirik momentini tanlash mumkin:

$$\mathcal{g}_{in}^0(X^{(n)}) = \frac{1}{n} \sum_{j=1}^n a_i(X_j), \quad i = \overline{1, s}.$$

1.1.1– Teorema. Faraz qilaylik, $g_i(\theta), i = \overline{1, s}$ funksiyalar Θ da uzluksiz hosilalarga ega bo'lib, $J_g(\theta) = \det \left\{ \left\| \frac{\partial g_i(\theta)}{\partial \theta_j} \right\|_{i,j=\overline{1,s}} \right\}$ - yakobian noldan farqli bo'lsin. Agar (1.1.1) sistema yechimi θ_n^0 yagona bo'lsa, u holda bu yechim θ uchun asosli baho bo'ladi.

Isboti. $G:\Theta \rightarrow \mathcal{H}$ desak, $G^{-1}:\mathcal{H} \rightarrow \Theta$ - bir qiymatli va uzluksizdir.

$\mathcal{G}_n^0 \xrightarrow[n \rightarrow \infty]{P} g_i, i = \overline{1, s}$, ekanidan, 1 ga yetarlicha yaqin ehtimollik bilan $\mathcal{G}_n^0 \in \mathcal{H}$. U holda (1.1.1) dan $\theta_n^0 = G^{-1}(\mathcal{G}_n^0)$ va G^{-1} ning uzluksiz ekanidan, $n \rightarrow \infty$ da $\theta_n^0 \xrightarrow{P} G^{-1}(G(\theta)) = 0$. ■

Momentlar usuli bahosi xossalari

1. Aytaylik $\theta^* = m^{-1}(\overline{g(x)})$ baho θ ni momentlar usulida topilgan bahosi bo'lsin. (bu yerda m^{-1} uzluksiz) U holda θ^* - asosli baho bo'ladi.

Isboti: Xinchinning katta sonlar qonuniga binoan

$$\overline{g(x)} = \frac{1}{n} \sum g(x) \xrightarrow{P} E_{\theta} g(x_i) = m(\theta),$$

m^{-1} uzluksizligidan

$$\theta^* = h^{-1}(\overline{g(x)}) \xrightarrow{P} h^{-1}(E_{\theta} g(x)) = h^{-1}(h(\theta)) = \theta.$$

2. Agar m funksiya θ nuqtada differensiallanuvchi, $\int g^2(x) P_{\theta} dx < \infty$ momentlar usulidagi baho asimptotik normal baho bo'ladi. Ushbu $((m'(\theta))^2, D_{\theta} g(x_1))$ parametrlar bilan.

Haqiqatga maksimal o'xshashlik usuli $\{P_{\theta}, \theta \in \Theta\} \ll \mu, f(x; \theta) = \frac{dP_{\theta}}{d\mu}(x)$

va $\theta_1 \neq \theta_2$ uchun $P_{\theta_1} \neq P_{\theta_2}, \theta_1, \theta_2 \in \Theta$ bo'lsin. Biz $\theta = (\theta_1, \theta_2, \dots, \theta_s) \in \Theta \subseteq R^{(s)}$ - vektor parametrni baholash masalasini qaraymiz. *Haqiqatga o'xshashlik funksiyasi* deb $\mathcal{B}^{(n)} \times \Theta$ da aniqlangan nomanfiy

$$f_n^*(x^{(n)}; \theta) = C f_n(x^{(n)}; \theta), (x^{(n)}; \theta) \in \mathcal{B}^{(n)} \times \Theta$$

ko'rinishdagi funksiyaga aytiladi. Bu yerda $C \in (0, \infty)$ - ko'paytuvchi θ ga bog'liq emas, ammo $x^{(n)}$ ga bog'liq bo'lishi mumkin va

$$f_n(x^{(n)}; \theta) = \prod_{i=1}^n f_n(x_i; \theta) - \text{tanlanmaning zichlik funksiyasi.}$$

1.1.1 – Ta'rif. *Haqiqatga maksimal o'xshashlik usuli bahosi.* (HMO'UB)

deb, quyidagi munosabatni qanoatlantiruvchi $\mathcal{B}^{(n)}$ - o'lchovli $\hat{\theta}_n(X^{(n)}): \mathcal{B}^{(n)} \rightarrow \Theta$ akslantirishga aytiladi:

$$f_n^*(X^{(n)}; \hat{\theta}_n) = \max_{\theta \in \Theta} \left\{ f_n^*(X^{(n)}; \theta) \right\}. \quad (1.1.2)$$

Demak, $\hat{\theta}_n$ ni topish, f_n^* ning maksimumini topishga ekvivalent masala ekan. f_n^* va $\ln f_n^*$ funksiyalar bir xil nuqtalarda ekstremumga erishishi sababli, (1.1.2) tenglikni $\ln f_n^*$ uchun ham yozish mumkin. Bu esa, o'z navbatida, amalda qulayliklarga olib keladi. U holda (1.1.2) tenglikni quyidagi ekvivalent ko'rinishda yozish mumkin:

$$\hat{\theta}_n(X^{(n)}) = \text{Arg max}_{\theta \in \Theta} \left\{ \int \ln f(x; \theta) \hat{P}_n(dx) \right\} = \text{Arg max}_{\theta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n \ln f(X_i; \theta) \right\}. \quad (1.1.3)$$

Ba'zi hollarda (1.1.2) tenglama yechimga ega bo'lmasligi ham mumkin. Odatda HMO'UB fiksirlangan $x^{(n)} \in \mathcal{B}^{(n)}$ da $f_n^*(x^{(n)}; \theta)$ - θ ning uzluksiz funksiyasi bo'lgan hollarda qo'llaniladi. HMO'UBLari yagona bo'lmasligi mumkin. Endi bahoning bunday nomlanishini biz faqat diskret holdagina (μ - sanoqli o'lchov) tushuntiramiz. Bu holda $f(x; \theta) = P_\theta(\xi = x)$ va

$$f_n(x^{(n)}; \theta) = P_\theta(X^{(n)} = x^{(n)}) = \prod_{i=1}^n P_\theta(\xi = x_i).$$

Demak, biz $\hat{\theta}_n$ sifatida f_n ehtimollikni maksimallashtiruvchi parametr qiymatini tanlar ekanmiz.

1. Agar $\Theta \subset R^{(s)}$ bo'lib, ixtiyoriy $x^{(n)} \in \mathcal{B}^{(n)}$ uchun $f_n^*(x^{(n)}; \theta)$ funksiya θ bo'yicha differensiallanuvchi va o'z maksimumiga Θ ning ichki nuqtasida (Θ ga biror oralig'i bilan tegishli bo'lgan nuqtada) erishsa, u holda $\hat{\theta}_n$ baho quyidagi shartni qanoatlantiradi:

$$\left. \frac{\partial f_n^*(x^{(n)}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0 \quad \text{yoki} \quad \left. \frac{\partial \ln f_n^*(x^{(n)}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0, \quad (1.1.4)$$

bu yerda

$$\frac{\partial \ln f_n}{\partial \theta} = \left(\frac{\partial \ln f_n}{\partial \theta_1}, \dots, \frac{\partial \ln f_n}{\partial \theta_s} \right).$$

2. Agar Kramer – Rao ma'nosida effektiv baho mavjud bo'lsa, uni HMO'UB yordamida topish mumkin.

3. Yana shuni ta'kidlab o'tamizki, agar HMO'UBsi $\hat{\theta}_n$ yagona bo'lsa, u yetarli statistika T ning funksiyasi bo'ladi.

$$\left. \frac{\partial \ln f_n(x^{(n)}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = \left. \frac{\partial \ln \Psi_n(T(x^{(n)}); \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0.$$

Ammo $\hat{\theta}_n$ ning o'zi yetarli statistika bo'lishi shart emas.

HMO'UBsining yana bir muhim xossalardan biri – uning parametrni almashtirishga nisbatan invariantligidir. Bu trivial da'voni isbotsiz keltiramiz. Θ - fazo $R^{(s)}$ dagi interval bo'lsin.

1.1.2– Teorema. (Invariantlik prinsipi (Zexna)). $g(\theta): \Theta \rightarrow \mathcal{H}$ funksiya berilgan bo'lib, $\mathcal{H}$ - fazo $R^{(k)}$ ($k \leq s$)dagi interval bo'lsin. Agar $\hat{\theta}_n$ baho θ parametr uchun HMO'UBsi bo'lsa, u holda $g(\theta)$ uchun $\hat{g}_n = g(\hat{\theta}_n)$ - HMO'UBsi bo'ladi.

1.2-§ Muhim taqsimotlar nomalum parametrlarining baholari va ularning hossalari

Gamma taqsimot va uning xossalari.

2.1.-Ta’rif. Agar ξ t.m. zichlik funksiyasi

$$f_{\xi}(x) = \begin{cases} \frac{\alpha^{\lambda}}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (1.2.1)$$

ko’rinishda bo’lsa, u holda ξ t.m. gamma taqsimotiga ega deyiladi, bu yerda $\alpha > 0$, $\lambda > 0$ va

$$\Gamma(\lambda) = \int_0^{\infty} t^{\lambda-1} e^{-t} dt \text{ - gamma funksiya:}$$

$$\Gamma(\lambda) = (\lambda - 1)\Gamma(\lambda - 1), \quad \Gamma(n) = (n - 1)!, \quad \Gamma(1/2) = \sqrt{\pi}.$$

Gamma taqsimotni $\Gamma_{\alpha, \lambda}$ orqali belgilaymiz.

$\xi : \Gamma_{\alpha, \lambda}$ t.m. xarakteristik funksiyasini hisoblaymiz:

$$\begin{aligned} \phi_{\xi}(t) &= M(e^{it\xi}) = \int_0^{\infty} e^{itx} \frac{\alpha^{\lambda}}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x} dx = \frac{\alpha^{\lambda}}{\Gamma(\lambda)} \int_0^{\infty} x^{\lambda-1} e^{-(\alpha - it)x} dx = \\ &= \frac{\alpha^{\lambda}}{\Gamma(\lambda)(\alpha - it)^{\lambda}} \int_0^{\infty} ((\alpha - it)x)^{\lambda-1} e^{-(\alpha - it)x} d(\alpha - it)x = \frac{\alpha^{\lambda}}{(\alpha - it)^{\lambda}} = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}. \end{aligned}$$

Demak, $\varphi_{\xi}(t) = Me^{it\xi} = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}$. Xarakteristik funksiya yordamida gamma

taqsimot momentlarini oson hisoblash mumkin: $M\xi = \frac{\lambda}{\alpha}, D\xi = \frac{\lambda}{\alpha^2}$.

Xossalari:

1) Agar $\xi_1, \dots, \xi_n$ bog'liqsiz t.m.lar bo'lib, $\xi_i : \Gamma_{\alpha, \lambda_i}, i=1, \dots, n$ bo'lsa, u holda $S_n = \sum_{j=1}^n \xi_j$ ning taqsimoti $\Gamma_{\alpha, \sum_{j=1}^n \lambda_j}$ bo'ladi.

Bu xossani isbotlash uchun xarakteristik funksiylardan foydalanamiz. $\Gamma_{\alpha, \lambda}$ taqsimotning xarakteristik funksiyasi $\varphi_{\xi}(t) = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}$ ga teng. Bog'liqsiz t.m.lar yig'indisining xarakteristik funksiyasi xarakteristik funksiylar ko'paytmasiga teng ekanligidan foydalansak, $S_n = \sum_{j=1}^n \xi_j$ t.m. xarakteristik funksiyasi

$$\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{\xi_j}(t) = \prod_{j=1}^n \left(1 - \frac{it}{\alpha}\right)^{-\lambda_j} = \left(1 - \frac{it}{\alpha}\right)^{-\sum_{j=1}^n \lambda_j}$$

bo'ladi. $\left(1 - \frac{it}{\alpha}\right)^{-\sum_{j=1}^n \lambda_j}$ xarakteristik funksiya esa $\Gamma_{\alpha, \sum_{j=1}^n \lambda_j}$ taqsimotning xarakteristik funksiyasidir. ■

2) Agar ξ standart normal taqsimotga ega bo'lsa, u holda ξ^2 tasodifiy miqdor $\Gamma_{1/2, 1/2}$ taqsimotga ega bo'ladi.

Buni ko'rsatish uchun avval ξ^2 t.m.ning taqsimotini topamiz. Agar $x \leq 0$ bo'lsa: $F_{\xi^2}(x) = P(\xi^2 < x) = 0$, $x > 0$ bo'lsa:

$$F_{\xi^2}(x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = F_{\xi}(\sqrt{x}) - F_{\xi}(-\sqrt{x})$$

bo'ladi. Bu yerda $F_{\xi}(x)$ - standart normal taqsimotning taqsimot funksiyasi. Endi ξ^2 t.m.ning zichlik funksiyasini topamiz. $x > 0$ da:

$$\begin{aligned} f_{\xi^2}(x) &= \left(F_{\xi^2}(x) \right)' = F_{\xi}'(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} + F_{\xi}'(-\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} = \\ &= \frac{1}{2\sqrt{x}} \left(f_{\xi}(\sqrt{x}) + f_{\xi}(-\sqrt{x}) \right) = \frac{1}{\sqrt{x}} \cdot f_{\xi}(\sqrt{x}) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}. \end{aligned}$$

$x \leq 0$ da: $f_{\xi^2}(x) = 0$.

Demak, ξ^2 t.m.ning zichlik funksiyasi

$$f_{\xi^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)} x^{1/2-1} e^{-x/2}, \quad x > 0,$$

$\Gamma_{1/2,1/2}$ taqsimot zichlik funksiyasiga teng ekan. ■

3) $\Gamma_{\alpha,1}$ taqsimot α parametrli ko'rsatkichli taqsimotdir.

Agar $\xi : \Gamma_{\alpha,1}$ bo'lsa, uning zichlik funksiyasi:

$$f_{\xi}(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

α parametrli ko'rsatkichli taqsimot zichlik funksiyasidir.

4) Agar ξ_1, ξ_2, K, ξ_k bog'liqsiz va standart normal taqsimotga ega t.m.lar bo'lsa, u holda $\eta = \xi_1^2 + \xi_2^2 + K + \xi_k^2 : \Gamma_{1/2,k/2}$ bo'ladi.

Momentlar usulida noma'lum parametrni baholash, eng sodda va keng qo'llaniladigan usullardan biridir. Gamma taqsimotining momentlar usulidagi bahosini topamiz. Gamma funksiyaning zichlik funksiyasi quyidagicha:

$$f(x; \theta_1; \theta_2) = \frac{x^{\theta_2-1} \cdot \exp(-x / \theta_2)}{\Gamma(\theta_2) \cdot \theta_1^{\theta_2}},$$

agar $x > 0$ bo'lsa.

Birinchi tartibli nazariy moment quyidagiga teng:

$$\begin{aligned} MX_1 &= \int_0^{\infty} \frac{x^{\theta_2} \cdot \exp(-x / \theta_2)}{\Gamma(\theta_2) \cdot \theta_1^{\theta_2}} dx = \frac{1}{\Gamma(\theta_2)} \int_0^{\infty} \left(\frac{x}{\theta_1} \right)^{\theta_2} e^{-\frac{x}{\theta_1}} dx = \left[\frac{x}{\theta_1} = t \right] = \\ &= \frac{\theta_1}{\Gamma(\theta_2)} \int_0^{\infty} t^{\theta_2} e^{-t} dt = -\frac{\theta_1}{\Gamma(\theta_2)} \int_0^{\infty} t^{\theta_2} de^{-t} = \\ &= -\frac{\theta_1}{\Gamma(\theta_2)} \left[t^{\theta_2} e^{-t} \Big|_0^{\infty} - \theta_2 \int_0^{\infty} t^{\theta_2-1} e^{-t} dt \right] = \theta_1 \theta_2 = \overline{X} \end{aligned}$$

Endi ikkinchi tartibli nazariy momentni hisoblaymiz:

$$\begin{aligned} MX_1^2 &= \int_0^{\infty} \frac{x^{\theta_2+1} \cdot \exp(-x / \theta_2)}{\Gamma(\theta_2) \cdot \theta_1^{\theta_2}} dx = \left[\frac{x}{\theta_1} = t \right] = \frac{\theta_1^2}{\Gamma(\theta_2)} \int_0^{\infty} t^{\theta_2+1} e^{-t} dt = \\ &= -\frac{\theta_1^2}{\Gamma(\theta_2)} \left[t^{\theta_2+1} e^{-t} \Big|_0^{\infty} - (\theta_2 + 1) \int_0^{\infty} t^{\theta_2} e^{-t} dt \right] = \frac{\theta_1^2 (\theta_2 + 1)}{\Gamma(\theta_2)} \int_0^{\infty} t^{\theta_2} e^{-t} dt = \\ &= -\frac{\theta_1^2 (\theta_2 + 1)}{\Gamma(\theta_2)} \left[t^{\theta_2} e^{-t} \Big|_0^{\infty} - \theta_2 \int_0^{\infty} t^{\theta_2-1} e^{-t} dt \right] = \theta_1^2 (\theta_2^2 + \theta_2) = \overline{X^2} \end{aligned}$$

Bu tenglamalardan foydalanib θ_1 va θ_2 larni topamiz.

$$\theta_1^2 = \frac{S^2}{\overline{X}} \text{ va } \theta_2^2 = \frac{(\overline{X})^2}{S^2}, \quad (1.2.2)$$

baho bo'lar ekan. Bu baholar momentlar usuli bahosi hossasiga ko'ra asosli baho bo'ladi.

Gamma taqsimot haqiqatga maksimal o`xshashlik bahosini topishimiz uchun biz, avvalambor, uning haqiqatga o`xshashlik funksiyasini tuzib olishimiz kerak bo`ladi. Gamma taqsimot haqiqatga o`xshashlik funksiyasi quyidagiga teng:

$$L(x, \theta) = \prod_{i=1}^n \frac{1}{\theta_1^{\theta_2} \Gamma(\theta_2)} \cdot x^{\theta_2-1} \cdot e^{-\frac{x}{\theta_1}} = \frac{1}{\theta_1^{n\theta_2} \Gamma^n(\theta_2)} \prod_{i=1}^n x_i^{\theta_2-1} \cdot e^{-\sum_{i=1}^n \frac{x_i}{\theta_1}}.$$

Haqiqatga o`xshashlik funksiyasini logarifmlaymiz:

$$l(x; \theta) = \ln L(x; \theta) = -n\theta_2 \ln \theta_1 - n \ln \Gamma(\theta_2) + (\theta_2 - 1) \ln \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{x_i}{\theta_1} \quad (1.2.3)$$

Yuqoridagi funksiyadan avval θ_1 bo`yicha keyin esa  $\theta_2$  bo`yicha hosila olib ularni 0 ga tenglab θ_1 va θ_2 lar uchun haqiqatga maksimal o`xshashlik usuli bahosini olamiz:

$$\frac{\partial l}{\partial \theta_1} = -\frac{n\theta_2}{\theta_1} + \frac{1}{\theta_1^2} \sum_{i=1}^n x_i = 0.$$

Bundan,

$$\theta_1 \theta_2 = \bar{X}, \quad (1.2.4)$$

ga teng ekanligi kelib chiqadi. Endi θ_2 bo`yicha hosila olamiz:

$$\frac{\partial l}{\partial \theta_2} = -n \ln \theta_1 - n \frac{\Gamma'(\theta_2)}{\Gamma(\theta_2)} + \sum_{i=1}^n \ln x_i = 0,$$

$\left[\frac{\Gamma'(\theta_2)}{\Gamma(\theta_2)} = \varphi(\theta_2) \right]$ deb belgilaymiz.

$$-n \ln \theta_1 - n \varphi(\theta_2) + \sum_{i=1}^n \ln x_i = 0$$

noma'lum parametrlarni bir tomonga qolganlarini esa ikkinchi tomonga o'tkazamiz:

$$\frac{1}{n} \sum_{i=1}^n \ln x_i = \ln \theta_1 + \varphi(\theta_2).$$

Bundan,

$$\sum_{i=1}^n \ln x_i^{\frac{1}{n}} = \ln \frac{\overline{X}}{\theta_2} + \varphi(\theta_2),$$

quyidagicha belgilash kiritamiz $\overline{X} \stackrel{\circ}{=} \prod_{i=1}^n x_i^{\frac{1}{n}}$, natijada tenglama quyidagi ko'rinishga kelib qoladi.

$$\ln \frac{\overline{X}}{X} = \varphi(\theta_2) - \ln \theta_2, \quad (1.2.5)$$

bu tenglamani yechishda sonli usullardan foydalanish kerak ana shunday usullardan biri Nyuton-Rafson usuli deb ataladi. Bu usul bilan biz keying paragrafda batafsil tanishamiz.

1.3-§. Nyuton-Rafson usuli

Nyuton-Rafson usuli Nyuton usuli nomi bilan ham mashhur bo`lib,

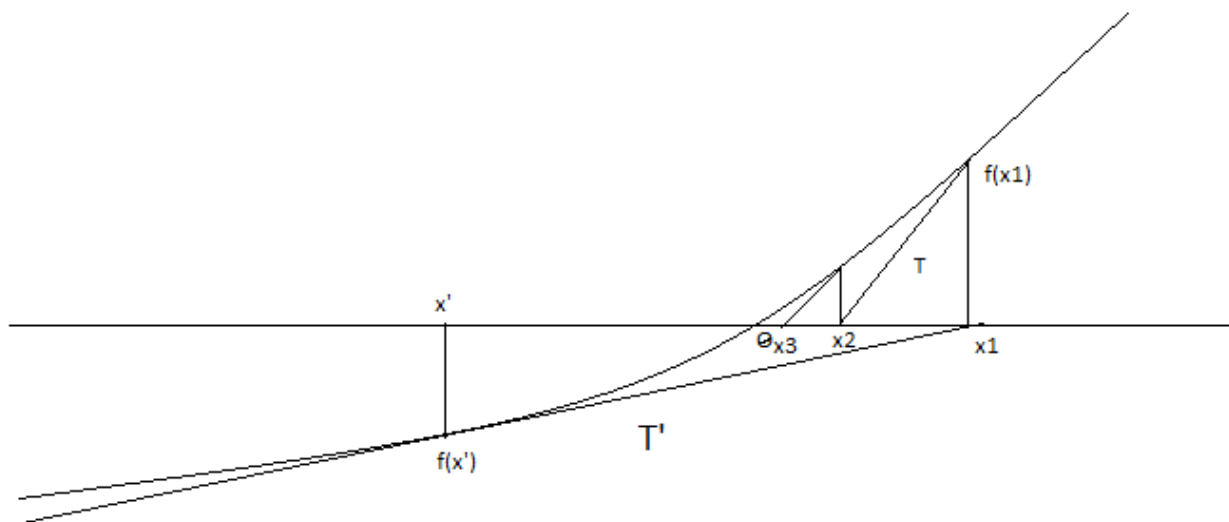
$$f(x) = 0, \quad (1.3.1)$$

tenglamaning yechimini topishda umumiy usul hisoblanadi.

Faraz qilaylik, $f(x)$ funksiya J intervalda ikki marta uzluksiz hosilaga ega bo`lib, uning biror  $\theta$  nuqtasida nolga teng bo`lsin:

$$f(\theta) = 0, \theta \in J.$$

Agar $x_1 \in J$ nuqta θ ga yaqin bo`lsa, u holda bu nuqtani dastlab Nyuton tomonidan taklif etilgan va keyinchalik Rafson tomonidan yaxshilangan usul yordamida aniqlash mumkin. Dastlab shu usulning geometrik ma`nosini ko`rib o`taylik (1-rasm). $f(x)$ funksiya grafigini θ nuqta atrofida ifodalaymiz:



1-rasm

$(x_1; f(x_1))$ nuqtadagi T urinma tenglamasi

$$\frac{T(x) - f(x_1)}{x - x_1} = f'(x_1). \quad (1.3.2)$$

T to'g'ri chiziq x o'qni $(x_2; 0)$ nuqtada kesib o'tadi, bu yerda $x_2 = x_1 - f(x_1)/f'(x_1)$ nuqtani (1) dan $x = x_2$, $T(x_2) = 0$ bo'lganida topamiz. Bu jarayonni davom ettirib, $x_1, x_2, \dots$ ketma-ketlikni aniqlaymiz. 1-rasmdan ko'rinadiki, bu ketma-ketlik θ nuqtaga monoton ravishda o'ngdan yaqinlashar ekan. Demak, bu ketma-ketlik biror $x^* \in [0; x_1]$ nuqtaga yaqinlashadi va

$$x_{n-1} - x_n = -\frac{f(x_n)}{f'(x_n)}, \quad (1.3.3)$$

tenglamadan

$$\lim_{n \rightarrow \infty} \frac{f(x_n)}{f'(x_n)} = 0, \text{ ya'ni } f(x^*) = 0.$$

Demak, $x^* = 0$ ekan. Biz masala yechimiga x_1 ni θ dan o'ng tomonda tanlaganimiz uchun erishdik. Agar 1-rasmda ko'rsatilganidek, jarayonni x' nuqtadan boshlaganimizda edi, u holda T' urinma $(x'; f(x'))$ nuqtadan o'tib x o'qi bilan kesishish nuqtasi θ dan x' ga nisbatan uzoqroqda bo'lib qolar edi.

Demak, $f(x)$ qanday bo'lganida $\{x_n\}$ ketma-ketlik θ ga yaqinlashadi degan savol tabiiydir. Quyidagi shartlar bu masala yechimini ta'minlaydi:

- a) x_{n+1} qiymati x_n va θ lar orasida yotadi;
- b) θ va x_1 lar orasida f ni nolga tenglashtiruvchi boshqa qiymat mavjud emas;

1-rasmda a) shart bajarilishi x_1 ning θ dan o'ngdaligi va f funksiya θ va x_1 orasida pastga qavariqligidan kelib chiqadi. x_1 ni to'g'ri tanlanishidan, $\text{sign}\{f(x_1)\} = +1$.

Yuqoridagilarni e'tiborga olgan holda quyidagi teoremani yoza olamiz. Unda Nyuton-Rafsonning $\{x_n\}$ ketma-ketligining θ ga monoton yaqinlashishining yetarli shartlari jamlangandir.

Teorema $f(x)$ funksiya $J = [\theta; x_1]$ kesmada differensiallanuvchi bo'lib, $f(\theta) = 0$ va $x \in J \setminus \{\theta\}$ uchun $f(x) \neq 0$ bo'lsin. Agar $f(x)$ funksiya J intervalda pastga qavariq va $f(x_1) > 0$ yoki $f(x)$ funksiya J da yuqoriga qavariq va $f(x_1) < 0$ bo'lsa, u holda Nyuton-Rafsonning $\{x_n\}$ ketma-ketligi θ ga monoton yaqinlashadi.

Bu teoremadagi shartlarni biroz qisqartirish natijasida Nyuton-Rafson usuli yaqinlashishining quyidagi Fyure shartlariga ega bo'lamiz.

Natija. Faraz qilaylik, $f(x)$ funksiya $J = [a; x_1]$ kesmada ikki marta differensiallanuvchi bo'lsin. Barcha $x \in J$ lar uchun $f'(x) \neq 0$ va $f''(x) \neq 0$ bo'lib, f'' hosila J da ishora saqlasin. Agar bundan tashqari, $\text{sign}\{f(x_1)\} = \text{sign}\{f''(x_1)\} = -\text{sign}\{f(a)\}$ bo'ladi va x_1 dan qurilgan $\{x_n\}$ Nyuton-Rafson ketma-ketligi θ ga monoton yaqinlashadi.

Ko'p hollarda haqiqatga maksimal o'xshashlik usuli tenglamasi

$$\frac{\partial \ln L_n(X^{(n)}; \theta)}{\partial \theta} = 0. \quad (1.3.4)$$

Noma'lum θ parametr ga nisbatan aniq yechimga ega bo'lmaydi. Bu yerda

$$L_n(X^{(n)}; \theta) = \prod_{k=1}^n f(X_k; \theta),$$

funksiya $X^{(n)} = (X_1, \dots, X_n)$ -statistik tanlanmaga nisbatan haqiqatga o'xshashlik funksiyasi. Bunga misol ko'raylik. Kuzatilayotgan ξ tasodifiy miqdor zichlik funksiyasi $f(x; \theta)$ bo'lgan gamma taqsimoti bo'lsin. Yuqorida ko'rganimizga ko'ra (1.3.4) tenglama quyidagiga teng:

$$\sum_{i=1}^n \ln x_i^{\frac{1}{n}} = \ln \frac{\overline{X}}{\theta_2} + \varphi(\theta_2).$$

Ko'rinib turibdiki yuqoridagi tenglama θ ga nisbatan aniq yechib bo'lmaydi. Shuning uchun biz Nyuton-Rafson usulidan foydalanamiz. Buning uchun $\hat{\theta}$ orqali (1.3.4) tenglama yechimi, ya'ni haqiqatga maksimal o'xshashlik bahosini belgilab olib, bu tenglama chap tomonidagi funksiyani $\hat{\theta}$ nuqta atrofida qatorga yoyamiz:

$$\begin{aligned} 0 = \frac{\partial}{\partial \theta} \ln L_n(X^{(n)}; \theta) &= \frac{\partial}{\partial \theta} \ln L_n(X^{(n)}; \theta_1) + \\ &+ (\theta - \theta_1) \frac{\partial^2}{\partial \theta^2} \ln L_n(X^{(n)}; \theta_1 + q(\theta - \theta_1)), \end{aligned} \quad (1.3.5)$$

bu yerda $0 \leq q \leq 1$ va θ_1 -boshlang'ich yechim. Agar (1.3.5) da $q = 0$ deb olsak, u holda $\hat{\theta}$ uchun 2-approximatsiyani olishimiz mumkin:

$$\theta_2 = \theta_1 - \frac{\frac{\partial}{\partial \theta} \ln L_n(X^{(n)}; \theta_1)}{\frac{\partial^2}{\partial \theta^2} \ln L_n(X^{(n)}; \theta_1)}. \quad (1.3.6)$$

Endi θ_1 o'rniga θ_2 ni qo'yib, (1.3.6) dan θ_3 ni aniqlaymiz va hokazo. Bu jarayonni davom ettirib, boshlang'ich yaqinlashish θ_1 dan $\{\theta_k, k \geq 1\}$ ketma-ketlikni hosil qilamiz:

$$\theta_{k+1} = \theta_k - \frac{\frac{\partial}{\partial \theta} \ln L_n(X^{(n)}; \theta_k)}{\frac{\partial^2}{\partial \theta^2} \ln L_n(X^{(n)}; \theta_k)}, \quad k = 1, 2, \dots \quad (1.3.7)$$

Agar (1.3.7) jarayonda boshlang'ich yechim θ_1 asosiy yechim $\hat{\theta}$ ga yaqin tanlangan bo'lib, $\frac{\partial^2}{\partial \theta^2} \ln L_n(X^{(n)}; \theta_k)$ $k = 1, 2, \dots$ musbat bo'lsa, u holda bu jarayon tezda yaqinlashadi.

II Bob. Noklassik modelda noma'lum parametrlarni baholash

2.1-§. To'liq bo'lmagan tanlanmalar modellari

Ta'rif: Tanlanma o'ng tomondan $x_{(2)}$ nuqtada 1-tur senzurlangan deyiladi, agar kuzatilayotgan $X_1, \dots, X_n$ tanlanma qiymatlari $M = (-\infty; x_{(2)}]$ to'plamga tegishli bo'lsa.

O'ng tomondan senzurlanishda tanlanma zichlik funksiyasi quyidagiga teng:

$$f(t) = f_{\beta}^c(t) \cdot S^{(1-t)}(t),$$

bu yerda: $S(t) = P(T > t) = 1 - F(t)$, $c_i = \begin{cases} 1 & , i - \text{qiymat senzurlanmagan,} \\ 0 & , i - \text{qiymat senzurlangan,} \end{cases}$

ga teng.

O'ng tomondan senzurlangan tanlanma haqiqatga o'xshashlik funksiyasi quyidagicha bo'ladi:

$$L(\beta) = \prod_{i=1}^n f_{\beta}^{c_i}(t_i) S^{(1-c_i)}(t_i).$$

Ta'rif: Tanlanma chap tomondan $x_{(1)}$ nuqtada 1-tur senzurlangan deyiladi, agar kuzatilayotgan $X_1, \dots, X_n$ tanlanma qiymatlari $M = [x_{(1)}; +\infty)$ to'plamga tegishli bo'lsa.

Chap tomondan senzurlanishda tanlanma zichlik funksiyasi quyidagiga teng:

$$f(t) = f_{\beta}^c(t) \cdot F^{(1-t)}(t).$$

Chap tomondan senzurlangan tanlanma haqiqatga o'xshashlik funksiyasi quyidagicha bo'ladi:

$$L(\beta) = \prod_{i=1}^n f_{\beta}^{c_i}(t_i) F^{(1-c_i)}(t_i).$$

2.2-§.Statistik baholashning EM algoritmi

Bundan keyingi masalalarda endi biz X vektorni to'liq tanlanma, Z vektorni senzurlangan tanlanma deb belgilaymiz.

$g_c(x; \theta)$ ni X tasodifiy miqdorning zichlik funksiyasi deb belgilaymiz. U holda X ning (agar u haqiqatdan kuzatilayotgan bo'lsa) haqiqatga o'xshashlik funksiyasi logarifmi quyidagicha bo'ladi:

$$\ln L_c(x) = \ln g_c(x; \theta). \quad (2.2.1)$$

EM algoritmi to'liq bo'lmagan tanlanmalarda haqiqatga o'xshashlik funksiyasi tenglamasini yechishda qo'llaniladi. Tanlanma to'liq bo'lmaganda

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0,$$

tenglamani yechimi aniq emas. Agarda X tasodifiy miqdor kuzatilmayotgan bo'lsa, u holda $\ln L_c(\theta)$ ham kuzatilmaydi va shuning uchun kuzatilayotgan Y tasodifiy miqdorning $\ln L_c(\theta)$ shartli matematik kutilmasidan foydalanamiz. EM algoritmining iteratsiyalarini qo'llaymiz:

E-qadam: $Q(\theta; \theta^{(k)})$ ni hisoblaymiz, bu yerda

$$Q(\theta; \theta^{(k)}) = E_{\theta^{(k)}}(\ln L_c(\theta) | Y). \quad (2.2.2)$$

M-qadam: $Q(\theta; \theta^{(k)})$ uchun keyingi $\theta^{(k+1)}$ maksimum nuqtani tanlaymiz:

$$Q(\theta^{(k+1)}; \theta^{(k)}) \geq Q(\theta; \theta^{(k)}). \quad (2.2.3)$$

Algoritmnining E va M qadamlari $\theta^{(k)}$ o'zining statsionar nuqtasiga erishmaguncha qaytadan amalga oshiriladi. Bu jarayon qachon to'xtatilishini $\|\theta^{(k+1)} - \theta^{(k)}\| < \varepsilon$ ayirma yordamida bilinadi, bu yerda ixtiyoriy $\varepsilon > 0$ son.

Agar $g_c(x; \theta)$ zichlik funksiya eksponensial oilaga tegishli bo'lsa:

$$g_c(x; \theta) = b(x) \exp(c(\theta)^T t(x)) / a(\theta). \quad (2.2.4)$$

EM algoritmining ikkala qadami ham odatda sodda ko'rinishda bo'ladi, bu yerda $t(x)$ vektor o'lchami $k \times 1 (k > d)$ bo'lgan yetarli statistika. $c(\theta)$ o'lchami $k \times 1$ vektor funksiya θ parametrdan olingan. $a(\theta)$ va $b(x)$ skalyar funksiya. Ko'p o'lchovli normal, Puasson, multinomial taqsimotlar eksponensial oilaga tegishli.

Eksponensial oilaga tegishli zichlik funksiya uchun algoritmnining E-qadami:

$$Q(\theta; \theta^{(k)}) = E_{\theta^{(k)}}(\ln b(X) | Y) + c(\theta)^T t^{(k)} - \ln a(\theta). \quad (2.2.5)$$

Bu yerda $t^{(k)} = E_{\theta^{(k)}}(t(X) | Y)$ yetarli statistikaning bahosi. M-qadamda $Q(\theta; \theta^{(k)})$ funksiyani maksimallashtirish zarur, lekin $E_{\theta^{(k)}}(\ln b(X) | Y)$ θ ga bog'liq emas shuning uchun algoritmnining iteratsiyasi soddalashadi:

E-qadam:

$$t^{(k)} = E_{\theta^{(k)}}(t(X) | Y), \quad (2.2.6)$$

hisoblanadi.

M-qadam:

$$\theta^{(k+1)} = \arg \max_{\theta \in \Theta} [c(\theta)^T t^{(k)} - \ln a(\theta)], \quad (2.2.7)$$

ni hisoblanadi.

X ning $Y=y$ dagi shartli zichlik funksiyasini $k(x|y; \theta) = g_c(x; \theta) / g(y; \theta)$ ko'rinishida belgilaymiz. U holda to'liq tanlanmaning haqiqatga o'xshashlik funksiyasi logarifmini quyidagicha bo'ladi:

$$\ln L_c(\theta) = \ln g_c(x; \theta) = \ln L(\theta) + \ln k(x | y; \theta). \quad (2.2.8)$$

$E_{\theta^{(k)}}(\cdot | Y)$ shartli matematik kutilmaning operatorini (2.2.8) tenglamaning ikkala tomoniga qo'llab

$$E_{\theta^{(k)}}(\ln k(X | Y; \theta) | y),$$

ni $H(\theta; \theta^{(k)})$ bilan belgilasak, quyidagi ifoda hosil bo'ladi:

$$Q(\theta; \theta^{(k)}) = \ln L(\theta) + H(\theta; \theta^{(k)}). \quad (2.2.9)$$

(2.2.9) formuladan quyidagi tenglikni keltirib chiqaramiz:

$$\begin{aligned} \ln L(\theta^{(k+1)}) - \ln L(\theta^{(k)}) &= [Q(\theta^{(k+1)}; \theta^{(k)}) - Q(\theta^{(k)}; \theta^{(k)})] - \\ &\quad - [H(\theta^{(k+1)}; \theta^{(k)}) - H(\theta^{(k)}; \theta^{(k)})]. \end{aligned} \quad (2.2.10)$$

Yensen tengsizligini qo'llab quyidagiga ega bo'lamiz:

$$H(\theta^{(k+1)}; \theta^{(k)}) \leq H(\theta^{(k)}; \theta^{(k)}).$$

Bizga (2.2.3) dan

$$Q(\theta^{(k+1)}; \theta^{(k)}) - Q(\theta; \theta^{(k)}) \geq 0,$$

ligi ma'lum. Shunday qilib haqiqatga o'xshashlik funksiyasi EM algoritmidan keyin kamaymaydi:

$$L(\theta^{(k+1)}) \geq L(\theta^{(k)}), \quad k=0,1,2,\dots \quad (2.2.11)$$

2.3-§. I tur senzurlanish modelida Gamma taqsimot nomalum parametrlarni baholash

(1.2.3) dan ma'lumki gamma taqsimot uchun $(\sum x_i, \sum \ln x_i)$ yetarli statistika bo'ladi. Biz yetarli statistikadan k va θ ning haqiqatga maksimal o'xshashlik bahosini hisoblashda foydalanamiz.

Agar tanlanma senzurlangan bo'lsa, EM algoritmi bizga haqiqatga maksimal o'xshashlik bahosini oladigan usulni beradi. EM algoritmi bizga to'liq bo'lmagan tanlanmaning haqiqatga o'xshashlik funksiyasini iteratsion ravishda maksimallashtirishimiz mumkin bo'lgan usulni taqdim etadi.

Bu holda biz yetarli statistikani, uning kutilmasini va tanlanmaning senzurlangan nuqtasini baholay olamiz. Bu statistika biz kuzatgan qiymatlarga va yetarli statistikaning kutilmasi parametrlarning hozirgi baholanishiga asoslangan. Bu EM algoritmining E qadami.

E- qadamda hisoblangan yetarli statistikaning baholaridan foydalanib biz parametrning yangi qiymatlarini hisoblashimiz mumkin, huddi (1.2.4) va (1.2.5) larorqali berilganidek bu EM algoritmining M qadamidir. Algoritmning E va M qadamlarini takrorlash orqali k va θ uchun yaxshiroq bahoga erishamiz.

Gamma taqsimot uchun biz yetarli statistikani va har bir yetarli statistikaning kutilmasini baholashimiz kerak. Bu statistika kuzatilgan qiymatlarga asoslangan.

Avvalambor, biz Gamma taqsimotini o'ng tomondan c bo'yicha senzurlangan deb qaraylik. Bu holda $(y_1, \dots, y_m, \dots, y_n)$ ko'rinishida berilgan tanlanmaning birinchi m taqiymati $(y_1, \dots, y_m)$ senzurlanmagan tanlanmaning qolgan $(n-m)$ tasi $(y_{m+1}, \dots, y_n)$ senzurlangan tanlanma hisoblanadi.

Bu holda $\sum_{i=1}^n x_i$ ni baholashda

$$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i + (n-m)E[x | \theta, k, x > c], \quad (2.3.1)$$

va $\sum_{i=1}^n \ln x_i$ ni baholashda:

$$\sum_{i=1}^n \ln x_i = \sum_{i=1}^n \ln y_i + (n-m)E[\ln(x) | \theta, k, x > c], \quad (2.3.2)$$

dan foydalanamiz. (2.3.1) dagi matematik kutilmani hisoblaymiz:

$$E[x | \theta, k, x > c] = \int_0^c x f(x | k, \theta, x > c) dx,$$

bu yerda

$$f(x | k, \theta, x > c) = \frac{\frac{1}{\theta^k \Gamma(k)} x^{k-1} \exp\left(-\frac{x}{\theta}\right)}{1 - F(c, k, \theta)}.$$

$F(c, k, \theta)$ gamma taqsimotning c nuqta bo'yicha taqsimot funksiyasidir.

$$\begin{aligned} E[x | x > c, k, \theta] &= \frac{\int_c^\infty x \frac{1}{\theta^k \Gamma(k)} \cdot x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx}{\int_c^\infty \frac{1}{\theta^k \Gamma(k)} \cdot x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx} = \\ &= \frac{\int_c^\infty x^{(1+k)-1} \exp\left(-\frac{x}{\theta}\right) dx}{\int_c^\infty x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx} = \left[u = \frac{x}{\theta} \Rightarrow du = \frac{dx}{\theta} \right], \end{aligned}$$

bundan,

$$E[x | x > c, k, \theta] = \frac{\int_{c/\theta}^\infty u^{(1+k)-1} \exp(-u) du}{\int_{c/\theta}^\infty u^{k-1} \exp(-u) du} = \theta \frac{\Gamma(k+1, c/\theta)}{\Gamma(k, c/\theta)},$$

bu yerda $\Gamma(k+1, c/\theta)$ senzurlangan gamma taqsimot.

Xuddi shunday (2.3.2) dagi matematik kutilma ham hisoblanadi.

$$\begin{aligned}
 E[\ln(x) | \theta, k, x > c] &= \int_0^c \ln(x) f(x | k, \theta, x > c) dx, \\
 E[\ln(x) | x > c, k, \theta] &= \frac{\int_c^\infty \ln(x) \frac{1}{\theta^k \Gamma(k)} \cdot x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx}{\int_c^\infty \frac{1}{\theta^k \Gamma(k)} \cdot x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx} = \\
 &= \frac{\int_c^\infty \ln(x) x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx}{\int_c^\infty x^{k-1} \exp\left(-\frac{x}{\theta}\right) dx} = \left[u = \frac{x}{\theta} \Rightarrow du = \frac{dx}{\theta} \right],
 \end{aligned}$$

deb belgilaymiz va

$$\begin{aligned}
 E[\ln(x) | x > c, k, \theta] &= \frac{\int_{c/\theta}^\infty \ln(u\theta) u^{k-1} \exp(-u) du}{\int_{c/\theta}^\infty u^{k-1} \exp(-u) du} = \\
 &= \frac{1}{\Gamma(k, c/\theta)} \int_{c/\theta}^\infty \ln(u\theta) u^{k-1} \exp(-u) du = \\
 &= \frac{1}{\Gamma(k, c/\theta)} \left(\ln(\theta) \int_{c/\theta}^\infty u^{k-1} \exp(-u) du + \int_{c/\theta}^\infty \ln(u) u^{k-1} \exp(-u) du \right) = \\
 &= \ln(\theta) + \frac{1}{\Gamma(k, c/\theta)} \int_{c/\theta}^\infty \frac{d}{dk} [u^{k-1} \exp(-u)] du = \\
 &= \ln(\theta) + \frac{1}{\Gamma(k, c/\theta)} \frac{d}{dk} \Gamma(k, c/\theta). \tag{2.3.4}
 \end{aligned}$$

Shu qiymatlarni yetarli statistika va (1.2.4), (1.2.5) tenglamalarga ishlatib biz parametrlarning baholarini yangilay olamiz. Keyin bu yangilangan parametrlar bilan jarayonni qaytarib, parametrlarni yana ham yaxshi baholarini olamiz. Bu

jarayonni yaqinlashish qoniqarli bo`lgunga qadar davom ettiramiz ya`ni, quyidagicha:

$$\begin{aligned}
 & -n\ln(\theta_{(t+1)}) - n\psi(k_{(t+1)}) + \sum_{i=1}^m \ln(y_i) + \\
 & + (n-m) \left(\ln \theta_{(t)} + \frac{1}{\Gamma(k_{(t)}, c / \theta_{(t)})} \frac{\partial}{\partial k} \Gamma(k_{(t)}, c / \theta_{(t)}) \right) = 0
 \end{aligned} \tag{2.3.5}$$

va

$$-\frac{nk_{(t+1)}}{\theta_{(t+1)}} + \frac{1}{\theta_{(t+1)}^2} \left(\sum_{i=1}^m y_i + (n-m)\theta_{(t)} \frac{\Gamma(k_{(t)} + 1, c / \theta_{(t)})}{\Gamma(k_{(t)}, c / \theta_{(t)})} \right) = 0.$$

Quyidagi jadvallarda parametrlari k, θ bo'lgan gamma taqsimotini noma'lum parametrlari uchun EM algoritmi yordamida baholar topilgan (bu yerda bosh to'plam $k = 2,5; \theta = 2$ parametrli gamma taqsimotga ega)

N=100

k uchun baho			
Senzurlanish	Senzuralanish foizi	o`rtachasi	xatolik
O`ng tomondan	10	2.5597	0.1339
	50	2.6206	0.2619
	80	2.8196	0.8295
Chap tomondan	10	2.5587	0.1438
	50	2.6429	0.3826
	80	3.0545	2.3225

θ uchun baho			
Senzurlanish	Senzuralanish foizi	o`rtachasi	xatolik
O`ng tomondan	10	1.9986	0.1011
	50	1.9978	0.2486
	80	2.0243	0.9516
Chap tomondan	10	1.9944	0.0974
	50	1.9845	0.1843
	80	1.9464	0.4660

N=1000

k uchun baho			
Senzurlanish	Senzuralanish foizi	o`rtachasi	xatolik
O`ng tomondan	10	2.5586	0.1355
	50	2.6133	0.2596
	80	2.799	0.848
Chap tomondan	10	2.5606	0.1447
	50	2.6471	0.397
	80	3.0385	2.2167

θ uchun baho			
Senzurlanish	Senzuralanish foizi	o`rtachasi	xatolik
O`ng tomondan	10	2.0001	0.1039
	50	2.0038	0.2465
	80	2.05	0.9759
Chap tomondan	10	1.9937	0.0994
	50	1.9845	0.1892
	80	1.9539	0.4731

N=10000

k uchun baho			
senzurlanish	Senzuralanish foizi	o`rtachasi	xatolik
O`ng tomondan	10	2.5010	$1.266 \cdot 10^{-3}$
	50	2.5009	$2.179 \cdot 10^{-3}$
	80	2.5023	$5.1 \cdot 10^{-3}$
Chap tomondan	10	2.5010	$1.326 \cdot 10^{-3}$
	50	2.5021	$3.163 \cdot 10^{-3}$
	80	2.5042	$1.131 \cdot 10^{-3}$

θ uchun baho			
Senzurlanish	Senzuralanish foizi	o`rtachasi	xatolik
O`ng tomondan	10	1.9994	$1.055 \cdot 10^{-3}$
	50	2.0001	$2.434 \cdot 10^{-3}$
	80	2.0009	$8.59 \cdot 10^{-3}$
Chap tomondan	10	1.9994	$9.952 \cdot 10^{-3}$
	50	1.9991	$1.877 \cdot 10^{-3}$
	80	1.9994	$4.827 \cdot 10^{-3}$

Xulosa

Ushbu bitiruv malakaviy ishda EM algoritmi yordamidada baho topish o`rganildi. Bunda umumiy holda Gamma taqsimoti nomalum parametri uchun EM algoritmi yordamida baho olindi. Amaliyotda yo`qotishga ega tizimlar va senzuralangan tizimlar uchun baho olish juda murakkab bo`lib, hisoblash jarayonini qiyinlashtiradi. Bunday vaziyatlarda EM algoritmi kabi iterativ usullar yordamida baho olish mumkin va kompyuter yordamida hisoblash jarayonida izlanayotgan baho olinadi. Ushbu bitiruv malakaviy ishda tanlanma elementlari 100ta 1000ta 10000ta; qiymatlar yo`qotilganligi kuchsiz (10%), o`rta(50%), kuchli(80) va senzurlanish o`ng va chap tomondan bo`lganda k va θ lar uchun baho olindi.

1-bobda baho va uning xossalari, ba`zi muhim taqsimot funksiyalar orasidagi bog`lanishlar va noma`lum parametrlari baholari, hamda Nyuton-Rafson usuli yoritilgan. Bitiruv malakaviy ishi 2- bobida senzurlagan tanlanmalarning ba`zi modellari, EM algoritmi chuqur o`rganilgan, ya`ni EM algoritmi yordamida gamma taqsimotining noma`lum parametrlari uchun baho topilgan.

Foydalanilgan adabiyotlar

1. А.А.Абдушукуров “ Статистика неполных наблюдений. Асимптотическая теория оценивания для неклассических моделей” «Университет», 2009. - с.
2. Боровков А.А. “Математическая статистика” – Учебник. – Институт математики 1997 г. 772 с. 3. Dempster A.P., Laird N.M. and Rubin D.B (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society B.39. pp1-23.
4. Kramer H. “Mathematical Methods of Statistics”, Princeton University Press. 1946. Y. 24 p.
5. Little R. J. , Rubin D.B. Statistical Analysis with Missing data Wiley, New York, 2002.
6. Nettleton D. Convergence properties of the EM algorithm in constrained parameter spaces Canad J. of Stat., 1999, 27.p. 639-648.
7. Shenton L.R. and Bowman K.O.(1977) Maximum likelihood estimation in small samples. Charles Griffen & Company ltd.
8. www.wikipedia.org.
9. www.teorver.ru.