معرفي MCMC

دكتر حجّت زراعتي : گروه اپيدميولوژي و آمار زيستي دانشگاه علوم پزشكي تهران، ۱۳۸۶

۱-۱. مقدمه

MCMC در اصل از زنجيرهاي ماركف براي انتگرال گيري مونت كارلويي استفاده مي نمايد. طرفداران آمار بيزي و گاهي اوقات پيروان آمار كلاسيک در بسياري از مواقع براي استنباط در مورد مدلهاي پارامتريک يا پيش بيني نيازمند محاسبه انتگرالهايي از توابع توزيع چند متغيره هستند. متخصصين آمار بيزي مايلند بر اساس داده هاي موجود از تابع توزيع پسين پارامترها انتگرال بگيرند و كلاسيک ها نيز لازم است كه از توزيع ها انتگرالگيري كنند. در انتگرال گيري مونت كارلويي نمونه هايي از توزيع مورد نظر انتخاب شده و سپس با محاسبه ميانگين نمونه تقريبي براي ميانگين واقعي بدست مي آورد. MCMC اين نمونه ها را با ايجاد يک زنجير ماركف در طول زماني طولاني اختيار مي كند. راههاي مختلفي براي توليد چنين زنجيرهايي وجود دارد، اما تمام اين روشها شامل نمونه گيري گيبز (Gibbs Sampler) و حالات خاصي از الگوريتم هاي متروپوليس(Metropolis) و هاستينگز(Hastings) هستند.

۱-۲. طرح موضوع

۱-۲-۱. استنباط بيزي

اكثر استفاده هايي كه تاكنون از MCMC شده است به استفاده آن در استنباط بيزي مرتبط بوده است. در يک چشم انداز بيزي، هيچ برتري اساسي بين پارامترهاي مدلهاي آماري و مشاهدات وجود ندارد، و همه آنها كميت هاي تصادفي هستند. فرض كنيد D دادة مشاهده شده، و θ نشان دهنده پارامتر توزيع و مقادير گمشده باشد. گام بعدي تعيين توزيع توأم $P(D,\theta)$ است. اين توزيع توأم به دو جزء توزيع پيشين $P(\theta)$ و توزيع شرطي $P(D|\theta)$ تفكيک مي شود، و داريم:

$$P(D,\theta) = P(D|\theta)P(\theta)$$

با داشتن مشاهده D، بنابر قضيه بيز مي توان توزيع θ به شرط D را بدست آورد:

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta}$$

اين توزيع را كه مبناي اصلي استنباط بيزي است، توزيع پسين θ مي نامند. اميد رياضي هر تابعي حقيقي از θ را مي توان بر مبناي توزيع پسين يافت:

$$E[f(\theta)|D] = \frac{\int f(\theta)P(\theta)P(D|\theta)d\theta}{\int P(\theta)P(D|\theta)d\theta}$$

اما محاسبه انتگرالهاي موجود در اين عبارت معمولا دشوار بوده(به ويژه براي بعدهاي بالا)، و در بسياري از موارد محاسبه تحليلي اين اميد رياضي غير ممكن است. راه جايگزين حل عددي مسئله خواهد بود، كه البته براي ابعاد بالاتر از ۲۰ دشوار و در برخي موارد نادرست خواهد بود. در چنين مواردي راه ديگر استفاده از تقريب لاپلاس و يا انتگرالگيري مونت كارلويي بر مبناي روشهاي MCMC است.

References:
Gilks W R, Richardson S, Spiegelhalter D J. *Introducing Markov Chain Monte Carlo, Chapman & Hall, New York, 1997.*

۱-۲-۲. محاسبه اميدهاي رياضي

محاسبه اميدهاي رياضي در توزيع هاي با بعد بالا موضوعي است که در استنباط کلاسيک نيز به آن پرداخته شده است. براي اجتناب از محدوديتهاي مباحث بيزي ما مسئله را در حالت کلي بررسي مي کنيم. فرض کنيد X برداري k مؤلفه اي از متغيرهاي تصادفي با توزيع $\pi(.)$ باشد. در بحثهاي بيزي، X شامل پارامترهاي مدل و داده هاي گم شده خواهد بود، و در روشهاي کلاسيک شامل داده ها يا آثار تصادفي(Random effect) مي باشد. در مدلهاي بيزي، $\pi(.)$ توزيع پسين خواهد بود، و در مدلهاي کلاسيک يک تابع درستنمايي است. بهرحال اميد رياضي مورد نظر عبارتست از :

$$E[f(X]=\frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \qquad (1.1)$$

که در آن f(.) تابع دلخواهي است. البته در حالتي که X گسسته باشد، انتگرال به سيگما تبديل خواهد شد.

۱-۳. MCMC

در اين قسمت مي خواهيم ببينيم چگونه مي توان اين اميد رياضي(رابطه 1.1) را با استفاده از روش MCMC محاسبه نمود. البته ابتدا جداگانه به دو مبحث انتگرال گيري مونت کارلويي و زنجيرهاي مارکف توجه مي نماييم، و سپس ترکيب آنها را مورد بررسي قرار خواهيم داد.

۱-۳-۱. انتگرال گيري مونت کارلويي

در اين روش $E[f(X)]$ با انتخاب دنباله نمونه هاي تصادفي $\{X_t, t=1,2,...,n\}$ از توزيع $\pi(.)$ و رابطه زير تقريب زده مي شود:

$$E[f(X)]\approx\frac{1}{n}\sum_{t=1}^{n} f(X_t)$$

يعني ميانگين جامعه با ميانگين نمونه برآورد مي شود. وقتي عناصر دنباله $\{X_t, t=1,2,...,n\}$ مستقل باشند، قانون اعداد بزرگ تأييد مي کند که با افزايش n دقت چنين برآوردي افزايش خواهد يافت. در حالت کلي انتخاب مستقل از هم عناصر $\{X_t\}$ از توزيع $\pi(.)$ امکانپذير نيست، زيرا در واقع $\pi(.)$ مي تواند غير استاندارد باشد. بهرحال لازم نيست $\{X_t\}$ مستقل باشد. $\{X_t\}$ عموما طبق فرآيندي توليد خواهد شد که با تسامح نمونه گيري از $\pi(.)$ به نسبت صحيح خوانده مي شود. يکي از راه هاي چنين اقدامي استفاده از زنجيرهاي مارکفي با توزيع مانا(Stationary) $\pi(.)$ است. نتيجه اين اقدام به نامگذاري MCMC موضوعيت مي بخشد.

References:
Gilks W R, Richardson S, Spiegelhalter D J. *Introducing Markov Chain Monte Carlo, Chapman & Hall, New York, 1997.*

۱-۳-۲. زنجيرهاي مارکف

دنباله متغيرهاي تصادفي $\{X_0, X_1, X_2, ...\}$ را در نظر بگيريد، بطوري که در هر زمان $t \geq 0$، حالت بعدي X_{t+1} از توزيع $P(X_{t+1}|X_t)$ و فقط وابسته به حالت کنوني يعني X_t انتخاب شود. در اينصورت با در دست داشتن اطلاعات X_t، حالت بعدي زنجير يعني X_{t+1} به تاريخچه قديمي تر زنجير $\{X_0, X_1, X_2, ..., X_{t-1}\}$ وابسته نخواهد بود. چنين دنباله اي يک زنجير مارکف ناميده شده، و $P(.|.)$ هسته انتقال (Transition Kernel) زنجير خوانده مي شود. در صورتيکه $P(.|.)$ به t بستگي نداشته باشد، زنجير از نظر زماني همگن (time- homogenous) ناميده مي شود.

سئوالي که وجود دارد اين است که حالت آغازين X_0 چگونه بر وقوع X_t اثر مي گذارد. اين سئوال به اين مفهوم است که توزيع X_t با اطلاع از X_0، که به صورت $P^{(t)}(X_t|X_0)$ نشان داده مي شود، چيست؟ در چنين حالتي به اطلاع از مداخلات متغيرهاي $\{X_1, X_2, ..., X_{t-1}\}$ نيازي نيست، بلکه X_t مستقيما" به X_0 وابسته خواهد بود. تحت شرايط منظم(Regularity Conditions)، زنجير به تدريج حالت ورودي را فراموش خواهد کرد و $P^{(t)}(.|X_0)$ سرانجام به يک توزيع ماناي يکتا(يا تغيير ناپذير) همگرا خواهد شد، که به t يا X_0 وابسته نخواهد بود. اين توزيع مانا را با $\Phi(.)$ نشان مي دهيم. در اينصورت با افزايش t، نقاط نمونه $\{X_t\}$ بطور افزاينده اي به نمونه هاي وابسته از توزيع $\Phi(.)$ ميل خواهند کرد.

بنابراين، پس از m مرحله تکرار(به اندازه کافي طولاني)، نقاط $\{X_t; t = m+1, m+2, ..., n\}$ نمونه هاي تقريبا وابسته از $\Phi(.)$ خواهند بود. فعلا با صرف نظر از پاسخ به پرسش چگونگي محاسبه m، به محاسبه برآورد اميد رياضي $E[f(X)]$ با استفاده از زنجيرهاي مارکف مي پردازيم، که در آن X داراي توزيع $\Phi(.)$ مي باشد. معمولا با صرف نظر از m نمونه غير لازم، برآوردگر زير محاسبه مي شود:

$$\bar{f} = \frac{1}{n-m}\sum_{t=m+1}^{n} f(X_t) \qquad (1.2)$$

اين برآورد گر را ميانگين ارگوديک(Ergodic Average) مي خوانند که تبديلي حافظ اندازه بوده و طبق قضيه ارگوديک به اميد رياضي مورد نظر همگرا است.

۱-۳-۳. الگوريتم متروپوليس_ هاستينگز

رابطه (1.2) نشان مي دهد که چگونه مي توان از يک زنجير مارکف براي برآورد $E[f(X)]$ روي فضاي توزيع $\Phi(.)$ استفاده نمود. به نظر مي رسد ما پاسخ مسئله خود را يافته ايم، اما ابتدا بايد بدانيم چگونه يک زنجير مارکف بر اساس توزيع ماناي $\Phi(.)$ که حاصل توزيع مورد نظر $\pi(.)$ است، بسازيم.

ساختن چنين زنجيرمارکفي ساده است. در الگوريتم متروپوليس- هاستينگز، در هر زمان t، حالت بعدي X_{t+1} اولين انتخاب نمونه اي از بين نقاط نامزد Y از يک توزيع پيشنهادي $q(.|X_t)$ خواهد بود. توزيع پيشنهادي ممکن است به حالت جاري X_t وابسته باشد. براي مثال، ممکن است $q(.|X_t)$ يک توزيع نرمال چند متغيره با ميانگين X و ماتريس کوواريانس معين باشد. نقطه نامزد Y با احتمال $\alpha(X_t, Y)$ پذيرفته خواهد شد، بطوريکه:

References:
Gilks W R, Richardson S, Spiegelhalter D J. *Introducing Markov Chain Monte Carlo, Chapman & Hall, New York, 1997.*

$$\alpha(X,Y) = \min\left(1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)}\right) \qquad (1.3)$$

اگر نقطه نامزد پذيرفته شود، حالت بعدي $X_{t+1} = Y$ خواهد بود. اگر نقطه نامزد رد شود، زنجير حركت نخواهد داشت يعني $X_{t+1} = X_t$. بنابر اين الگوريتم متروپوليس- هاستينگز بسيار ساده است:

```
Initialize X0; set   t=0.
 Repeat{
      Sample a point Y from q(.|Xt)
      Sample a Uniform(0,1) random variable U
      If  U ≤ α(Xt,Y)   set  Xt+1 = Y
            Otherwise set  Xt+1 = Xt
       Increment  t
 }.
```

بايستي به خاطر داشت كه توزيع پيشنهادي $q(.|.)$ مي تواند هر شكل دلخواهي داشته باشد، و توزيع ماناي زنجير همان توزيع $\pi(.)$ خواهد بود.

References:
Gilks W R, Richardson S, Spiegelhalter D J. *Introducing Markov Chain Monte Carlo, Chapman & Hall, New York, 1997.*