Rao-Blackwellization

Note Title 4/12/2006

Two points to this lecture:

(1.) How to improve the efficiency of a statistical estimator by adding more statistics.

(2.) When should you sample? And when use analytic techniques?

Recall efficiency,
no. of samples depends on variance of estimator.

(Page 2) How Many Independent Samples are Needed?

Want to estimate $\mu = \sum_x h(x)\pi(x)$

– eg. The expected cost to build a pipeline in Sicily.

Take m samples $x^{(1)}, \ldots, x^{(m)}$ from $\pi(x)$

Estimate. $S_m = \frac{1}{m}\sum_{i=1}^{m} h(x^{(i)})$

The expectation of the estimate.

$$E_\pi[S_m] = \sum_{x^{(1)},\ldots x^{(m)}} \frac{1}{m}\sum_{i=1}^{m} h(x^{(i)}) \prod_{i=1}^{m} \pi(x^{(i)})$$

$$E_\pi[S_m] = \sum_x h(x)\pi(x) = \mu \text{ , unbiased estimate.}$$

The variance of the estimate

$$Var_\pi[S_m] = E_\pi\{S_m^2\} - \{E_\pi\{S_m\}\}^2$$

$$Var_\pi[S_m] = \sum_{\{x^{(i)}:\, i=1 \text{ to } m\}} \prod_{k=1}^{m} \pi(x_k)\, \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} h(x^{(i)})\,h(x^{(j)}) - \mu^2$$

$$Var_\pi[S_m] = \frac{1}{m^2}\cdot m\, E_\pi\{h^2(x)\} + \frac{1}{m^2}(m^2-m)\mu^2 - \mu^2.$$

$$Var_\pi[S_m] = 1/m\ \bar\sigma^2$$

where $\bar\sigma^2 := E_\pi\{(h(x)-\bar h)^2\}$, $\bar h := E_\pi\{h(x)\}$

MacKay says 12 samples. – cost is likely to differ from μ by $\pm\bar\sigma$ anyway. (~ Central Limit Theorem).

Page 3

One View of Rao-Blackwell.

An initial estimator $\alpha(x)$

Improve it by adding statistics $T(x)$ to get a new estimator

$$\alpha_1(x) = E(\alpha(x) | T(x))$$

expectation.

Claim: α_1 is always better than α (unless it is the same).

Technically, $E((\alpha_1(x) - \theta)^2) \leq E((\alpha(x) - \theta)^2)$

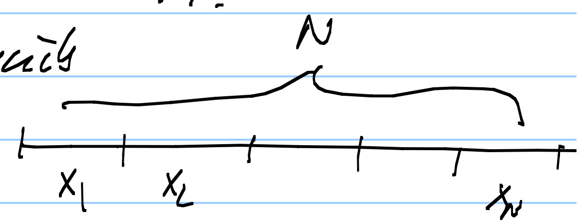
(Proof later in lecture.)

E.G. Data generated by Poisson process.

$$P(M \text{ events in time } N) = \frac{e^{-\lambda N}(\lambda N)^M}{M!}$$

<u>Divide</u> N into N time units

Let X_i be the no. of counts in the i^{th} time window.



Task: estimate λ

Data $\underline{X} = (X_1, X_2, \ldots, X_N)$

Page 4. Initial Estimator

$\alpha(\underline{x}) = \begin{cases} 1, & \text{if } x_1 = 0 \\ 0, & \text{otherwise.} \end{cases}$ // Ignores $x_2, x_3, \ldots$ etc.)

In period 1, $P(X_1 \text{ counts}) = \dfrac{e^{-\lambda}\lambda^{x_1}}{x_1!}$

$$\sum_x \alpha(x) P(X_1 \text{ counts}) = e^{-\lambda} \quad (= P(X_1 = 0)).$$

So $\alpha(x)$ is an unbiased estimator of $e^{-\lambda}$ (i.e. $E[\alpha(x)] = e^{-\lambda}$), but the variance is high.

$$\sum_{x_1} (\delta(x_1) - e^{-\lambda})^2 \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} = e^{-\lambda} - e^{-2\lambda}$$

Better to use statistic: $T(\underline{x}) = x_1 + \ldots + x_N$.

$\alpha_1(\underline{x}) = E(\alpha \mid x_1 + \ldots + x_N)$

This is the probability that we have a total of $x_1 + \ldots + x_N$ counts, but $x_1 = 0$. (i.e. all the counts are in the remaining $n-1$ time periods.)

This equals $\left(1 - \frac{1}{n}\right)^{x_1 + \ldots + x_n}$

If n is suff large

$$x_1 + \ldots + x_n \approx n\lambda$$

$$\left(1 - \frac{1}{n}\right)^{x_1 + \ldots + x_n} \approx \left(1 - \frac{1}{n}\right)^{n\lambda} \approx e^{-\lambda}$$

So we estimate $e^{-\lambda}$ well from one sample $\underline{x}$

(or take $\frac{1}{n}\{\alpha(x_1) + \alpha(x_2) + \ldots + \alpha(x_n)\}$)

Note Title 4/2/2006

General Rao-Blackwellization.

(Page 5.) Rule of Thumb → when sampling, do as much as possible analytically.

Goal: Estimate $E_{\pi}[h(\underline{x})] = \int_D h(\underline{x})\pi(\underline{x})d\underline{x}$

Monte-Carlo says sample from $\pi(\underline{x})$ to get i.i.d. $\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots, \underline{x}^{(m)}$

$$\hat{I}_m = \frac{1}{m}\sum_{i=1}^{m} h(\underline{x}^{(i)}) \qquad \lim_{m\to\infty} \hat{I}_m = I$$

$$\frac{1}{\sqrt{m}}(\hat{I}_m - I) \sim N(0, \sigma^2)$$

$$\sigma = \mathrm{Var}_{\pi}(h)$$

No. samples required $\sim \sqrt{m}/\sigma$.

(e.g. $X = \{X/\pi(X), \pi(X)\}$)

Suppose we can decompose $\underline{X}$ into two parts $\underline{X} = (\underline{X}_1, \underline{X}_2)$ | $\pi_2(\underline{X}_2) = \sum_{\underline{x}_1} \pi(\underline{x}_1, \underline{x}_2)$.

<u>and</u> $E[h(\underline{x}) | \underline{x}_2] = \sum_{\underline{x}_1} h(\underline{x}_1, \underline{x}_2)\pi(\underline{x}_1 | \underline{x}_2)$ can be computed analytically.

New estimator

$$\tilde{I}_m = \frac{1}{m}\sum_{i=1}^{m} E[h(\underline{x}) | \underline{x}_2^{(i)}]$$

where $\underline{x}_2^{(1)}, \ldots \underline{x}_2^{(m)}$ are i.i.d. samples from $\pi_2(\underline{x}_2)$

Page 2. We have two estimates $\hat{I}_m$ and $\hat{\hat{I}}_m$ for I. which is better?

Both estimates are unbiased — the estimates are r.v.'s depending on the samples $\underline{x}^{(1)}, \dots \underline{x}^{(m)}$ from distribution $\pi(\underline{x})$.

$$\hat{I}_m(\underline{x}^{(1)}, \dots \underline{x}^{(m)}).$$

Its expectation is

$$E_\pi[\hat{I}_m] = \sum_{\underline{x}^{(1)}, \dots \underline{x}^{(m)}} \hat{I}_m(\underline{x}^{(1)}, \dots, \underline{x}^{(m)})\, \pi(\underline{x}^{(1)}) \dots \pi(\underline{x}^{(m)})$$

$$= \sum_{\underline{x}^{(1)}, \dots, \underline{x}^{(m)}} \frac{1}{m} \sum_{i=1}^{m} h(\underline{x}^i)\, \pi(\underline{x}^{(1)}) \dots \pi(\underline{x}^{(m)})$$

$$= \sum_{\underline{x}} h(\underline{x})\pi(\underline{x}) = I.$$ Hence unbiased.

Similarly, it can be checked that

$$E_\pi[\hat{\hat{I}}_m] = \sum_{\underline{x}_2} \sum_{\underline{x}_1} E[h(\underline{x}) \mid \underline{x}_2)\pi(\underline{x}_2).$$

$$= \sum_{\underline{x}_1, \underline{x}_2} h(\underline{x}_1, \underline{x}_2)\pi(\underline{x}_1 \mid \underline{x}_2)\pi(\underline{x}_2)$$

$$= I.$$

Page 3.

Example. X_1, X_2

suppose $\pi(x_1 \mid x_2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x_1-x_2)^2/2\sigma^2}$

$\pi(x_2) =$ uniform distribution $0 \le x_2 \le 1$.

$h(x_1, x_2) = (x_1^2 + x_2^2)$

Then $E[h(x_1,x_2) \mid x_2] = \int_{-\infty}^{\infty} (x_1^2+x_2^2) \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x_1-x_2)^2/2\sigma^2} dx_1$

$$= (x_2^2+\sigma^2) + x_2^2 = \sigma^2 + 2x_2^2.$$

Hence

$$\bar{\bar{I}}_m = \sigma^2 + \frac{2}{m} \sum_{i=1}^{m} (x_2^{(i)})^2$$

where $\{x_2^{(i)} : i = 1 \text{ to } m\}$ are iid sample from $\pi(x_2) \sim$ uniform distribution.

Note: $\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu^2 + \sigma^2.$

Because $\int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$ (definition)

Expand $(x-\mu)^2 = x^2 - 2x\mu + \mu^2$ — result follows

Page 4.

How many samples are needed?

Recall that $\sqrt{m}\,(\bar{I}_m - I) \sim N(0, \sigma^2)$

$$\sigma^2 = Var_{\pi}(h(x)).$$

<u>Claim.</u>

$$var_{\pi}\{h(\underline{x})\} = var_{\pi(x_2)}\{E_{\pi(x_1|x_2)}\{h(\underline{x})\,|\,x_2\}\} + E_{\pi(x_2)}\{var_{\pi(x_1|x_2)}\{h(\underline{x})\,|\,x_2\}\}.$$

This implies that

$$var\{h(\underline{x})\} \geq var\{E\{h(\underline{x})\,|\,x_2\}\}$$

with equality only if $E\{var\{h(\underline{x})\,|\,x_2\}\} = 0$

only if $h(\underline{x})$ is a deterministic function of x_2.

<u>Proof of Claim.</u>

$$var\{h(\underline{x})\} = \sum_{x_1,x_2}\{h(x_1,x_2)\}^2\pi(x_1,x_2) - \Big(\sum_{x_1,x_2} h(x_1,x_2)\pi(x_1,x_2)\Big)^2$$

$$= \sum_{x_2}\Big\{\sum_{x_1}\{h(x_1,x_2)\}^2\pi(x_1|x_2) - \Big(\sum_{x_1} h(x_1,x_2)\pi(x_1|x_2)\Big)^2 + \Big(\sum_{x_1} h(x_1,x_2)\pi(x_1|x_2)\Big)^2\Big\}\pi(x_2) - \Big\{\sum_{x_2}\sum_{x_1} h(x_1,x_2)\pi(x_1|x_2)\pi(x_2)\Big\}^2$$

$$= E_{\pi(x_2)}\{var_{\pi(x_1|x_2)}(h(\underline{x})|x_2)\} + Var_{\pi(x_2)}\{E_{\pi(x_1|x_2)}\{h(\underline{x})\}\}$$