

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 $\mu = \sum h(x)\pi(x)$

- eg. The expected cost to build a pipeline in Sicily (Mackey reading)

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

Estimator $S_m = \frac{1}{m} \sum_{i=1}^m h(x^{(i)})$. This is a random variable which depends on the samples.

The expectation of the estimator is

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

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

$$= \frac{1}{m} \sum_{i=1}^m \sum_{x^{(i)}} h(x^{(i)}) \pi(x^{(i)}) = \frac{1}{m} \cdot m \sum_x h(x) \pi(x) = \sum_x h(x) \pi(x)$$

(using $\sum_{x^{(j)}} \pi(x^{(j)}) = 1$, for all j)

The variance of the estimator is:

$$\text{Var}_{\pi}[S_m] = E_{\pi}\{S_m^2\} - \left(E_{\pi}[S_m]\right)^2$$

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

$$\text{Var}_{\pi}[S_m] = \frac{1}{m^2} \cdot m E_{\pi}\{h^2(x)\} + \frac{1}{m^2} (m^2 - m) \mu^2 - \mu^2$$

So $\text{Var}_{\pi}[S_m] = \frac{1}{m} \bar{\sigma}^2$

$\underbrace{\hspace{10em}}_{\text{diagonal terms } i=j}$
 $\underbrace{\hspace{10em}}_{\text{off-diagonal terms } i \neq j}$

where $\bar{\sigma}^2 = E_{\pi}[(h(x) - \bar{h})^2]$, with $\bar{h} = E_{\pi}\{h(x)\} = \sum_x \pi(x) h(x)$.

Page 3

One View of Rao-Blackwell.

An initial estimator $d(x)$

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

$$d_1(x) = E[d(x) | T(x)]$$

≔ expectation.

Claim: d_1 is always better than d (unless it is the same).

Technically, $E[(d_1(x) - \theta)^2] \leq E[(d(x) - \theta)^2]$

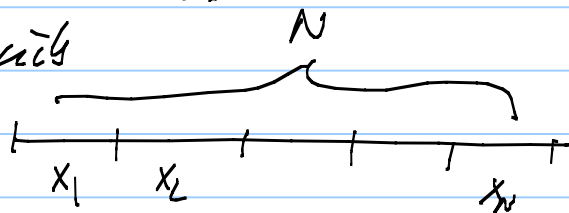
(Proof later in lecture.)

E.G. Data generated by Poisson process.

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

Divide N into N time units

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



Task: estimate λ

Data $\underline{x} = (x_1, x_2, \dots, x_n)$

Page 4.

Initial Estimator

$$d(\underline{x}) = \begin{cases} 1, & \text{if } x_1 = 0 \\ 0, & \text{otherwise.} \end{cases}$$

// ignores x_2, x_3, \dots
etc.)

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

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

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

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

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

$$d_1(\underline{x}) = E[d \mid x_1 + \dots + x_n]$$

This is the probability that we have a total of $x_1 + \dots + x_n$ counts, but $x_1 = 0$. (i.e. all events are in the

This equals $\left(1 - \frac{1}{n}\right)^{x_1 + \dots + x_n}$ remaining $n-1$ time periods.

If n is suff large

$$x_1 + \dots + x_n \approx n\lambda \rightarrow \text{expectation wrt. the Poisson Process}$$

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

So we estimate $e^{-\lambda}$ well from one sample \underline{x}
(or take $\frac{1}{n} \{d(x_1) + d(x_2) + \dots + d(x_n)\}$)

General Rao-Blackwellization.

Note Title

4/2/2006

Page 5)

Rule of Thumb \rightarrow When sampling, do as much as possible analytically.

$$\text{Goal: Estimate } I = E_{\pi} [h(\underline{x})] = \int_{\mathcal{D}} h(\underline{x}) \pi(\underline{x}) d\underline{x}$$

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

$$\bar{I}_m = \frac{1}{m} \sum_{i=1}^m h(\underline{x}^{(i)}) \quad \lim_{m \rightarrow \infty} \bar{I}_m = I$$

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

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

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

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

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

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

New estimator $\bar{I}_m = \frac{1}{m} \sum_{i=1}^m E[h(\underline{x}) \mid \underline{x}_2^{(i)}]$

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

Page 6.

We have two estimates I_m and \widehat{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 $\pi(\underline{x})$.

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

Its expectation is

$$\begin{aligned} E_{\pi}[\widehat{I}_m] &= \sum_{\underline{x}^{(1)}, \dots, \underline{x}^{(m)}} \widehat{I}_m(\underline{x}^{(1)}, \dots, \underline{x}^{(m)}) \pi(\underline{x}^{(1)}) \dots \pi(\underline{x}^{(m)}) \\ &= \sum_{\underline{x}^{(1)}, \dots, \underline{x}^{(m)}} \frac{1}{m} \sum_{i=1}^m h(\underline{x}^{(i)}) \pi(\underline{x}^{(1)}) \dots \pi(\underline{x}^{(m)}) \\ &= \sum_{\underline{x}} h(\underline{x}) \pi(\underline{x}) = I. \quad \text{Hence unbiased.} \end{aligned}$$

Similarly, it can be checked that

$$\begin{aligned} E_{\pi}[\widehat{I}_m] &= \sum_{\underline{x}_2} \sum_{\underline{x}_1} E[h(\underline{x}) | \underline{x}_2] \pi(\underline{x}_2) \\ &= \sum_{\underline{x}_1, \underline{x}_2} h(\underline{x}_1, \underline{x}_2) \pi(\underline{x}_1 | \underline{x}_2) \pi(\underline{x}_2) \\ &= I. \end{aligned}$$

Page 7.

Example

X_1, X_2

suppose

$$\pi(x_1 | x_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - x_2)^2}{2\sigma^2}}$$

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

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

$$\begin{aligned} \text{Then } E[h(x_1, x_2) | x_2] &= \int_{-\infty}^{\infty} (x_1^2 + x_2^2) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1 - x_2)^2}{2\sigma^2}} dx_1 \\ &= (x_2^2 + \sigma^2) + x_2^2 = \sigma^2 + 2x_2^2. \end{aligned}$$

Hence

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

where $\{x_2^{(i)}; i=1 \dots m\}$ are iid samples from $\pi(x_2) \sim$ uniform distribution.

Note: For this example we can compute

$$I = \int_0^1 (\sigma^2 + 2x^2) dx = \sigma^2 + 2/3 \text{ analytically}$$

Even better!

Note:

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

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

$$\text{Expand } (x-\mu)^2 = x^2 - 2x\mu + \mu^2 - \text{result follows}$$

Page 8.

How many samples are needed?

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

$$\sigma^2 = \text{Var}_{\pi} \{h(x)\}.$$

Claim.
$$\text{Var}_{\pi} \{h(x)\} = \text{Var}_{\pi(x_2)} \left\{ \text{Var}_{\pi(x_1|x_2)} \{h(x)\} \right\} + E_{\pi(x_2)} \left\{ \text{Var}_{\pi(x_1|x_2)} \{h(x)\} \right\}$$

This implies that

$$\text{Var}_{\pi} \{h(x)\} \geq \text{Var}_{\pi(x_2)} \left\{ E_{\pi(x_1|x_2)} \{h(x)\} \right\}$$

with equality only if $E_{\pi(x_2)} \left\{ \text{Var}_{\pi(x_1|x_2)} \{h(x)\} \right\} = 0$

i.e. only if $h(x)$ is a deterministic function of x_2 .

Proof of Claim.

$$\begin{aligned} \text{Var} \{h(x)\} &= \sum_{x_1, x_2} \left\{ h(x_1, x_2) \right\}^2 \pi(x_1, x_2) - \left(\sum_{x_1, x_2} h(x_1, x_2) \pi(x_1, x_2) \right)^2 \\ &= \sum_{x_2} \left\{ \sum_{x_1} \left\{ h(x_1, x_2) \right\}^2 \pi(x_1|x_2) - \left(\sum_{x_1} h(x_1, x_2) \pi(x_1|x_2) \right)^2 \right. \\ &\quad \left. + \left(\sum_{x_1} h(x_1, x_2) \pi(x_1|x_2) \right)^2 \pi(x_2) - \left(\sum_{x_2} \sum_{x_1} h(x_1, x_2) \pi(x_1|x_2) \pi(x_2) \right)^2 \right\} \\ &= E_{\pi(x_2)} \left\{ \text{Var}_{\pi(x_1|x_2)} \{h(x)\} \right\} + \text{Var}_{\pi(x_2)} \left\{ E_{\pi(x_1|x_2)} \{h(x)\} \right\} \end{aligned}$$