به نام خدا

دانشکده مهندسی کامپیوتر

فرایندهای تصادفی
نیم‌سال اول ۱۴۰۳-۱۴۰۲
حمیدرضا ربیعی

پایان‌ترم

۱ کوتاه پاسخ (۱۵ نمره)

۱. **۵ نمره،** آیا می‌توان تنها با داشتن ماتریس گذار[1] یک زنجیره‌ی مارکوف، نادوره‌ای[2] بودن آن را تعیین کرد؟ در صورت پاسخ مثبت الگوریتمی برای این کار ارائه دهید، در غیر این صورت دلیل این امر را بیان کنید.

۲. **۵ نمره،** یک زنجیره‌ی مارکوف با تعداد حالات نامحدود و شمارا در نظر بگیرید.[3] در صورتی که بازای هر دو حالت، احتمال رسیدن از حالت اول به حالت دوم در تعداد گام‌های محدود، مثبت باشد. زنجیره‌ی مارکوف حتما بازگشتی[4] است.

۳. **۵ نمره،** اگر Y یک تخمین‌گر unbiased برای پارامتر θ باشد، آنگاه e^Y یک تخمین‌گر unbiased برای e^θ است.

[1] Transition Matrix
[2] Aperiodic
[3] Countably Infinite
[4] Recurrent

پاسخ

(b) The same as above, yes. We need to calculate all powers of transition matrix from 2 to $2(n-1)+n-2$. The last one, ensures us that the longest possible path from a state to another state, looping over all the states other than the mentioned two, and returning to initial state has been considered (it might be even less than that but the question does not ask for the most efficient algorithm)! Then, $\text{period}(i) = \text{GCD}([\forall t, T_{ii}^{t} > 0])$. If all periods are 1, then the chain is apreiodic.

The statement is **False**. In a Markov chain with countably infinite states, even if every state has a positive probability of reaching any other state in a finite number of steps, the chain may still be transient rather than recurrent. Transient states have a non-zero probability of never returning to the initial state. The key distinction lies in whether, on average, the chain will eventually return or not.

د:

$$E[Y] = \theta$$

مثال نقض $X \sim \text{uniform}(0, \theta) \longrightarrow Y = 2X$

$$E_x[2x] = \int_0^\theta 2x \cdot \frac{1}{\theta}\, dx = \frac{2}{\theta} \cdot \frac{1}{2} x^2 \Big|_0^\theta = \theta$$

$$E_x[e^{2x}] = \int_0^\theta e^{2x} \cdot \frac{1}{\theta}\, dx = \frac{1}{\theta} \cdot \frac{1}{2} e^{2x} \Big|_0^\theta$$

$$= \frac{1}{2\theta} e^{2\theta} - \frac{1}{2\theta} = \frac{1}{2\theta}(e^{2\theta} - 1) \neq e^{\theta}$$

۲ سوالات چهار گزینه‌ای (۲۰ نمره)

۱. **۵ نمره**، اگر $s(f)$ چگالی طیف توانی[5] یک فرایند تصادفی WSS باشد، کدام یک از گزاره‌های زیر همواره درست است؟

(آ) $s(0) \geq s(f)$

(ب) $s(f) \geq 0$

(ج) $s(-f) = -s(f)$

(د) $\int_{-\infty}^{\infty} s(f)df = 0$

۲. **۵ نمره**، کدام گزینه از ویژگی‌های تابع خودهمبستگی[6] $R_X(\tau)$ فرایند تصادفی X_t است؟

[5] Power Spectral Density
[6] Auto-correlation Function

(آ) تابعی قطعی[۷] است که بیشینه‌ی آن در $\tau = 0$ رخ می‌دهد.
(ب) تابعی قطعی و متناوب است.
(ج) یک فرایند تصادفی ایستا است.
(د) یک فرایند تصادفی متناوب است.

۳. **۵ نمره،** $X(t)$ را یک فرایند WSS در نظر بگیرید که آن را به عنوان ورودی به سیستم LTI با پاسخ ضربه‌ی $h(t) = \delta(t+a) - \delta(t-a)$ می‌دهیم. خرجی این سیستم را فرایند $Y(t)$ بگیرید. تابع خودهمبستگی $R_{YY}(\tau)$ بر حسب تابع خودهمبستگی $R_{XX}(\tau)$ در کدام یک از گزینه‌های زیر آمده است.

(آ) $R_{YY}(\tau) = 2R_{XX}(\tau) - R_{XX}(\tau + 2a) - R_{XX}(\tau - 2a)$
(ب) $R_{YY}(\tau) = 2R_{XX}(\tau) + R_{XX}(\tau + a) + R_{XX}(\tau - a)$
(ج) $R_{YY}(\tau) = 2R_{XX}(\tau) + R_{XX}(\tau + 2a) + R_{XX}(\tau - 2a)$
(د) $R_{YY}(\tau) = 2R_{XX}(\tau) - R_{XX}(\tau + a) - R_{XX}(\tau - a)$

۴. **۵ نمره،** کدام یک از گزینه‌های زیر از ویژگی‌های فرایند تصادفی گاوسی نیست؟

(آ) یک فرایند گاوسی بوسیله‌ی میانگین و خودهمبستگی به صورت کامل تعریف می‌شود.
(ب) اگر یک فرایند گاوسی WSS باشد آن‌گاه SSS هم هست.
(ج) اگر یک فرایند گاوسی به عنوان ورودی به یک سامانه‌ی LTI ورودی داده شود، خروجی سیستم هم یک فرایند گاوسی خواهد بود.
(د) اگر دو فرایند که به صورت توام گاوسی هستند، ناهمبسته[۸] باشند، آن‌گاه آن‌ها از به صورت آماری به هم وابسته هستند.

[7] Deterministic
[8] Uncorrelated

پاسخ

For a random process X(t), the **Fourier transform of the autocorrelation function $R_X(\tau)$,** denoted by $S_X(f)$, is called the **power spectral density** of process X(t).

Thus $S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) \exp^{-j2\pi f\tau} \, d\tau$

Some Properties of Power spectral Density:

→ $S_X(f)$ is a **real-valued even function**, and **$S_X(f) \geq 0$ for all f.**

→The integral over the frequency range - π to π is proportional to the variance of a zero-mean random process and 2π is the proportionality coefficient.

→ $R_X(0) = \int_{-\infty}^{\infty} S_X(f) df$

$R_X(0)$ represents the **average power** in a WSS process.

A random process is said to be **strictly stationary** if, for each **n**, and each choice of $t_1, t_2, \ldots, t_n$, the **joint CDF** of $X(t_1), X(t_2), \ldots, X(t_n)$ is the same as the joint CDF of $X(t_1 + t), X(t_2 + t), \ldots, X(t_n + t)$, for any **t**. That is, **the statistics of the random process are invariant to time shifts.**

Strict sense stationarity is a very strong condition to require on a random process. In practice, it is enough if **only first and second-order conditions are satisfied**.

Important Points

Wide Sense Stationarity:

A random process X(t) is said to be wide sense stationary (WSS) if **its mean does not change with time** and its **autocorrelation function depends on only the time difference between the samples.**

mX(t) = mX and **$R_X(t + \tau, t) = R_X(\tau)$**

Additional Information

Some Properties of the Autocorrelation Function:

1) $R_X(\tau)$ is an **even function** of τ .

2) $R_X(0) \geq 0$. (Since the second moment of X(t) is ≥ 0.)

3) R_X **archives its maximum absolute value at 0**, i.e., $|R_X(\tau)| \leq R_X(0)$ (by the Cauchy-Schwarz inequality).

Concept:

If a WSS random process **X(t)** passing through an LTI system with **Y(t)** output random process provided impulse response **h(t).**

Then

$\mathbf{R_{yy}(\tau) = R_{xx}(\tau) \otimes h(\tau) \otimes h(-\tau)}$

Where,

$R_{yy}(\tau)$ → autocorrelation function of the output random process Y(t)

$R_{xx}(\tau)$ → autocorrelation function of random process X(t)

Calculation:

Given;

$h(t) = \delta(t + a) - \delta(t - a)$

$R_{yy}(\tau) = R_{xx}(\tau) \otimes h(\tau) \otimes h(-\tau)$

$\Rightarrow R_{yy}(\tau) = R_{xx}(\tau) \otimes [h(\tau) \otimes h(-\tau)]$

$\Rightarrow R_{yy}(\tau) = R_{xx}(\tau) \otimes [\{\delta(\tau + a) - \delta(\tau - a)\} \otimes \{\delta(-\tau + a) - \delta(-\tau - a)\}]$

$\Rightarrow R_{yy}(\tau) = R_{xx}(\tau) \otimes [\{\delta(\tau + a) - \delta(\tau - a)\} \otimes \{\delta(\tau - a) - \delta(\tau + a)\}]$ $[\because \delta(at + b) = (1/|a|)\delta(t + b/a)]$

$\Rightarrow R_{yy}(\tau) = R_{xx}(\tau) \otimes [2\delta(\tau) - \delta(\tau + 2a)\} - \delta(\tau - a) - \delta(\tau - 2a)]$

$\therefore$ $\mathbf{R_{yy}(\tau) = 2R_{xx}(\tau) - R_{xx}(\tau + 2a) - R_{xx}(\tau - 2a)}$ $[\because f(x) \otimes \delta(t + a) = f(t + a)]$

Option 4 is not a property of a Gaussian random process.

Gaussian random process:

In probability theory and statistics, a Gaussian process is a stochastic process, such that every finite collection of those random variables has a multivariate normal distribution, i.e. every finite linear combination of them is normally distributed.

The distribution of a Gaussian process is the joint distribution of all those (infinitely many) random variables, and as such, it is a distribution over functions with a continuous domain, e.g. time or space.

Properties of a Gaussian random process:

1. A Gaussian process is completely characterized by its mean and autocorrelation function.
2. If a Gaussian process is a wide-sense stationery, then it is stationary in the strict sense too.
3. If a Gaussian process is given as input to an LTI system, the output process is also Gaussian.
4. **If two processes which are jointly Gaussian are uncorrelated, then they are not statistically dependent.**

۳ پوآسون (۲۰ نمره)

تعداد زیادی تفنگ داریم که همه آنها در صورت فعال بودن با هر یک از رویدادهای یک فرآیند پوآسون $X(t)$ با نرخ λ یک گلوله شلیک می‌کنند. در ابتدا همه تفنگ‌ها غیر از تفنگ اول را غیرفعال می‌کنیم. با هر شلیک گلوله توسط تفنگ i ، تفنگ $i+1$ فعال می‌شود. با توجه به اطلاعات داده شده به سوالات زیر پاسخ دهید.

۱. **۱۰ نمره،** الف. میانگین تعداد شلیک‌ها در بازه $[0, t]$ چقدر است؟

۲. **۱۰ نمره،** ب. میانگین تعداد شلیک‌ها در بازه $[t_1, t_2]$ چقدر است؟

پاسخ

N_t را تعداد گلوله‌های شلیک شده در بازه $[0, t]$ تعریف می‌کنیم. داریم

$$\mathbb{E}[N_t|X_t] = \frac{(x_t)(x_t+1)}{2} = \frac{x_t^2 + x_t}{2}$$

همینطور میدانیم که

$$\mathbb{E}[X_t^2] = \lambda t + (\lambda t)^2$$
$$\mathbb{E}[N_t] = \lambda t + \frac{(\lambda t)^2}{2}$$
$$\mathbb{E}[N_{t_2} - N_{t_1}] = \frac{(\lambda t_2)^2 - (\lambda t_1)^2}{2} + \lambda(t_2 - t_1)$$

۴ تخمین‌گر بیشینه‌ی درستنمایی[۹] (۲۵ نمره)

متغیرهای تصادفی $Y_1, Y_2, ..., Y_n$ را به صورت زیر در نظر بگیرید.

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, ..., n$$

در رابطه‌ی بالا $x_1, x_2, ..., x_n$ را مقادیر ثابت در نظر بگیرید، همچنین فرض کنید $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. σ^2 را دانسته در نظر بگیرید.

۱. **۵ نمره،** آماره‌ی کافی دو بعدی برای (β, σ^2) بدست آورید.

۲. **۵ نمره،** تخمین‌گر بیشینه‌ی درستمایی را برای β بدست آورید. همچنین نشان دهید که تخمین‌گر بدست آمده unbiased است.

۳. **۵ نمره،** توزیع پاسخ بدست آمده در بخش دوم را پیدا کنید.

۴. **۵ نمره،** نشان دهید مقدار $\sum Y_i / \sum x_i$ یک تخمین‌گر unbiased برای β است.

۵. **۵ نمره،** واریانس $\sum Y_i / \sum x_i$ را محاسبه کرده و آن را با واریانس تخمین‌گر بیشینه‌ی درستنمایی مقایسه کنید.

[9] Maximum Likelihood Estimator

پاسخ

a.

$$
\begin{aligned}
L(\theta|\mathbf{y}) &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2)\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\beta^2 \sum_i x_i^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i\right).
\end{aligned}
$$

By Theorem 6.1.2, $(\sum_i Y_i^2, \sum_i x_i Y_i)$ is a sufficient statistic for (β, σ^2).

b.

$$\log L(\beta,\sigma^2|\mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i - \frac{\beta^2}{2\sigma^2}\sum_i x_i^2.$$

For a fixed value of σ^2,

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2}\sum_i x_i y_i - \frac{\beta}{\sigma^2}\sum_i x_i^2 \overset{\text{set}}{=} 0 \quad \Rightarrow \quad \hat\beta = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$

Also,

$$\frac{\partial^2 \log L}{\partial \beta^2} = \frac{1}{\sigma^2}\sum_i x_i^2 < 0,$$

so it is a maximum. Because $\hat\beta$ does not depend on σ^2, it is the MLE. And $\hat\beta$ is unbiased because

$$\mathrm{E}\,\hat\beta = \frac{\sum_i x_i \mathrm{E}\,Y_i}{\sum_i x_i^2} = \frac{\sum_i x_i \cdot \beta x_i}{\sum_i x_i^2} = \beta.$$

c. $\hat\beta = \sum_i a_i Y_i$, where $a_i = x_i / \sum_j x_j^2$ are constants. By Corollary 4.6.10, $\hat\beta$ is normally distributed with mean β, and

$$\mathrm{Var}\,\hat\beta = \sum_i a_i^2 \mathrm{Var}\,Y_i = \sum_i \left(\frac{x_i}{\sum_j x_j^2}\right)^2 \sigma^2 = \frac{\sum_i x_i^2}{(\sum_j x_j^2)^2}\sigma^2 = \frac{\sigma^2}{\sum_i x_i^2}.$$

$$\mathrm{E}\frac{\sum_i Y_i}{\sum_i x_i} = \frac{1}{\sum_i x_i}\sum_i \mathrm{E}\,Y_i = \frac{1}{\sum_i x_i}\sum_i \beta x_i = \beta.$$

$$\mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right) = \frac{1}{(\sum_i x_i)^2}\sum_i \mathrm{Var}\,Y_i = \frac{\sum_i \sigma^2}{(\sum_i x_i)^2} = \frac{n\sigma^2}{n^2\bar x^2} = \frac{\sigma^2}{n\bar x^2}.$$

Because $\sum_i x_i^2 - n\bar x^2 = \sum_i (x_i - \bar x)^2 \ge 0$, $\sum_i x_i^2 \ge n\bar x^2$. Hence,

$$\mathrm{Var}\,\hat\beta = \frac{\sigma^2}{\sum_i x_i^2} \le \frac{\sigma^2}{n\bar x^2} = \mathrm{Var}\left(\frac{\sum_i Y_i}{\sum_i x_i}\right).$$

۵ زنجیره‌ی مارکوف (۱۵ نمره)

فرض کنید دو جعبه داریم که مجموعا N توپ درون آن‌ها قرار دارد. در هر گام از زنجیره‌ی مارکوف یک توپ به صورت کاملا تصادفی و یکنواخت انتخاب شده و به جعبه‌ی دیگر منتقل می‌شود. فرض کنید X_n تعداد توپ جعبه‌ی اول در زمان n باشد. در

این صورت ماترین گذار این زنجیره‌ی مارکوف به صورت زیر تعریف می‌شود:

$$P(x,y) = \begin{cases} \frac{N-x}{N}, & \text{if } y = x+1 \\ \frac{x}{N}, & \text{if } y = x-1 \\ 0, & \text{otherwise.} \end{cases}$$

نشان دهید که توزیع ایستای یکتا برای این زنجیره‌ی مارکوف یک توزیع binomial با پارامترهای N و $\frac{1}{2}$ است.

پاسخ

Solution. Notice that any of two states in the chain communicates each other. Indeed, if $i, j \in \mathcal{S} = \{0, \cdots, N\}$ are distinct states, then

$$\rho_{ij} \geq \begin{cases} P(i,i+1)P(i+1,i+2)\cdots P(j-1,j) > 0, & \text{if } i < j, \\ P(i,i-1)P(i-1,i-2)\cdots P(j+1,j) > 0, & \text{if } i > j. \end{cases}$$

So the chain is finite and irreducible, and there exists a unique stationary distribution. Then by this uniqueness, it suffices to show that a binomial distribution with parameters N and $1/2$ is a stationary distribution.

- **1st Solution.** Let $\pi(x) = \binom{N}{x}\frac{1}{2^N}$. We check that π solves $\pi = \pi P$. Indeed, for $0 < x < N$,

$$\begin{aligned}(\pi P)_x &= \pi(x-1)P(x-1,x) + \pi(x+1)P(x+1,x) \\ &= \binom{N}{x-1}\frac{1}{2^N}\cdot\frac{N-x+1}{N} + \binom{N}{x+1}\frac{1}{2^N}\cdot\frac{x+1}{N} \\ &= \left\{\binom{N-1}{x-1} + \binom{N-1}{x}\right\}\frac{1}{2^N} \\ &= \binom{N}{x}\frac{1}{2^N} = \pi(x).\end{aligned}$$

  Similar computation shows that this equality is also true when $x = 0$ and N. Therefore π is a stationary distribution.

- **2nd Solution.** We attempt to find a stationary distribution out of P. By the way the chain is constructed, it is not unreasonable to seek for a distribution π that satisfies the detailed balanced condition. Since $P(x,y) > 0$ only when $|x-y| = 1$, this boils down to solving the equation

$$\forall x \in \{1, \cdots, N\} : \quad \pi(x-1)\frac{N-x+1}{N} = \pi(x)\frac{x}{N}.$$

۶ نمونه برداری مبتنی بر اهمیت[10] (۲۵ نمره)

در این سوال به دنبال تخمین امیدریاضی تابع کران‌دار $0 < m < f(x) < M$ روی اندازه‌ی احتمال[11] $p(x)$ هستیم. فرض کنید یک اندازه‌ی احتمال دیگر $q(x)$ را در اختیار داریم که می‌توانیم نمونه‌برداری را از آن انجام دهیم. همچنین فرض کنید بازای تمامی x هایی که $p(x) > 0$ است، داشته باشیم $q(x) > 0$ که این خاصیت را با $p << q$ نشان می‌دهیم.
فرض کنید نمونه‌های مستقل $x_1, x_2, \ldots, x_n$ از q نمونه‌برداری شده‌اند. تخمین‌گرهای زیر را برای تخمین $\mu = \mathbb{E}_p[f(x)]$ در نظر بگیرید.

$$\hat{\mu}_{\text{IPS}} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)}$$

$$\hat{\mu}_{\text{NIPS}} = \frac{\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)}}{\sum_{i=1}^{n}\frac{p(x_i)}{q(x_i)}}$$

[10] Importance Sampling
[11] Probability Measure

$$\hat{\mu}_{\text{PM}} = \frac{1}{n} f(x_i) \frac{2p(x_i)}{p(x_i) + q(x_i)}$$

۱. **۵ نمره،** ابتدا نشان دهید $\hat{\mu}_{\text{IPS}}$ یک تخمین‌گر unbiased است. سپس نشان دهید در صورتی که $p = q$ باشد تمامی تخمین‌گرها unbiased هستند.

۲. **۵ نمره،** فرض کنید $\mu > 0$. با داشتن $n \leq N$، $\mu > \tau$ و $\delta > 0$ ثابت کنید توزیع $q(x) >> p(x)$ وجود دارد به طوری که بازای هر $n \leq N$:

$$\mathbb{P}(|\hat{\mu}_{\text{IPS}} - \mu| \geq \tau) \geq 1 - \delta$$

همچنین ثابت کنید برای $n \leq N$ با انتخاب مناسب p و q میتوان واریانس $\hat{\mu}_{\text{IPS}}$ را هر اندازه بزرگ کرد.

۳. **۵ نمره،** ثابت کنید $\hat{\mu}_{\text{IPS}}$ به صورت مجانبی unbiased است. یعنی

$$\lim_{n \to \infty} \hat{\mu}_{\text{IPS}} = \mu$$

همچنین نشان دهید به ازای همه‌ی nها:

$$\mathbb{V}(\hat{\mu}_{\text{IPS}}) \leq \frac{(M-m)^2}{4}$$

۴. **۵ نمره** فرض کنید $f(x) = 1$ و داشته باشیم:

$$\mathbb{P}(p(x) \geq \alpha q(x)) \geq \frac{1}{1+\alpha}$$

به ازای $\alpha > 6$. ثابت کنید:

$$\mathbb{E}[\hat{\mu}_{\text{PM}}] - \mu \geq \frac{1}{4}$$

همچنین ثابت کنید

$$\mathbb{V}[\hat{\mu}_{\text{PM}}] - \mu \leq \frac{(M-m)^2}{n}$$

۵. **۵ نمره،** تخمین‌گرهای معرفی شده را با یکدیگر مقایسه کنید، مزایا و معایب هر یک را بررسی کرده و مشخص کنید هر کدام در چه شرایطی بر دیگران ترجیح داده می‌شوند.

پاسخ

۱. Trivial.

۲. Set $A = [t, +\infty)$ where t is chosen such that:

$$\mathbb{P}_p(x > t) \leq \epsilon$$

We set values of q on A^c such that $q(A^c) = \epsilon'$ and define q on A as:

$$q(x) = kp(x)$$
$$k = \frac{1-\epsilon'}{\int_A p(x)dx} \geq \frac{1-\epsilon'}{\epsilon}$$

in all are they if So .A^c in are x_i the of none that $(1-\epsilon')^N$ least at of probability a is there Now
have: we ،A

$$\hat{\mu}_{IPS} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{kp(x_i)} \leq \frac{1}{n}\sum_{i=1}^{n} M\frac{\epsilon}{1-\epsilon'} = M\frac{\epsilon}{1-\epsilon'}$$

easily is which solved، is problem we ،$M\frac{\epsilon}{1-\epsilon'} \leq \mu - \tau$ and $(1-\epsilon')^N \geq 1-\delta$ have we if So
achievable.
variance، For

$$\mathbb{V}(\hat{\mu}_{IPS}) = \frac{1}{n}\left(\mathbb{E}_q\left[f(x)\frac{p^2(x)}{q^2(x)}\right] - \mu^2\right)$$

set، we if Again

$$t : \mathbb{P}_q(X \geq t) = \epsilon$$

and

$$p(x) = \frac{1}{2\epsilon}q(x)$$

variance، on bound lower following the have will We .$x < t$ for arbitrary and $x \geq t$ for

$$\mathbb{V}(\hat{\mu}_{IPS}) \geq \frac{m}{4n\epsilon}$$

large. arbitrarily set can Which

۳. have: We

$$\hat{\mu}_{NIPS} = \frac{\frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)}}{\frac{1}{n}\sum_{i=1}^{n}\frac{p(x_i)}{q(x_i)}}$$

we numbers large of law strong the by ،$B = \frac{1}{n}\sum_{i=1}^{n}\frac{p(x_i)}{q(x_i)}$ and $A = \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)}$ set we if
have:

$$A \to \mathbb{E}_q\left[f(x)\frac{p(x)}{q(x)}\right] = \mu$$
$$B \to \mathbb{E}_q\left[\frac{p(x)}{q(x)}\right] = 1$$

have: we So

$$\hat{\mu}_{NIPS} = \frac{A}{B} \to \frac{\mu}{1} = \mu$$

have، we variance، the For

$$m = \frac{1}{n}\sum_{i=1}^{n} m\frac{p(x_i)}{q(x_i)} \leq \hat{\mu}_{IPS} \leq \frac{1}{n}\sum_{i=1}^{n} M\frac{p(x_i)}{q(x_i)} = M$$

variance. its on bound proposed the have we So .$[m, M]$ in bounded is $\hat{\mu}_{IPS}$ So have، We .۴

$$\begin{aligned}\mathbb{E}[\hat{\mu}_{PM}] - \mu &= \mathbb{E}[\hat{\mu}_{PM}] - \mathbb{E}[\hat{\mu}_{IPS}] = \mathbb{E}[\hat{\mu}_{PM} - \hat{\mu}_{IPS}] \\ &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{2p(x_i)}{p(x_i)+q(x_i)} - \frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{q(x_i)}\right] \\ &= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} f(x_i)\left(\frac{2p(x_i)}{p(x_i)+q(x_i)} - \frac{p(x_i)}{q(x_i)}\right)\right] \\ &= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} f(x_i)\frac{p(x_i)}{p(x_i)+q(x_i)}\left(\frac{p(x_i)}{q(x_i)} - 1\right)\right]\end{aligned}$$

،$Y = \frac{p(X)}{q(X)}$ set we if

$$\begin{aligned}\mathbb{E}[\hat{\mu}_{PM}] - \mu = \mathbb{E}\left[\frac{x(x-1)}{x+1}\right] &= \mathbb{E}\left[\frac{x(x-1)}{x+1}|x<1\right]P(x<1) + \mathbb{E}\left[\frac{x(x-1)}{x+1}|x>1\right]P(x>1) \\ &\geq -\frac{1}{4} + \mathbb{E}\left[\frac{x(x-1)}{x+1}|x>1\right]P(x>1) \geq -\frac{1}{4} + \mathbb{E}\left[\frac{x(x-1)}{x+1}|x>\alpha\right]P(x>\alpha) \\ &\geq -\frac{1}{4} + \frac{\alpha(\alpha-1)}{\alpha+1}\frac{1}{1+\alpha}\end{aligned}$$

have، we so ،$\frac{\alpha(\alpha-1)}{(\alpha+1)} \geq \frac{1}{2}$ ،$\alpha > 6$ for have We

$$\mathbb{E}[\hat{\mu}_{PM}] - \mu \geq -\frac{1}{4} + \frac{1}{2} = \frac{1}{4}$$

have، we variance، For

$$\mathbb{V}(\hat{\mu}_{PM}) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i)\frac{2p(x_i)}{p(x_i)+q(x_i)}\right) = \frac{1}{n}\mathbb{V}\left(f(X)\frac{2p(X)}{p(X)+q(X)}\right) = \frac{1}{n}\mathbb{V}(g(X))$$

.$\mathbb{V}(g(X)) \leq (M-m)^2$ have we So .$2m \leq g(X) \leq 2M$ have: we but

۵. be can bias the samples، of number fixed a for but unbiased، is estimator IPS seen، have we As
the as decays which high، arbitrarily be can it and variance high a has also It large. arbitrarily
know we or samples of number large very a have we when So increases. samples of number
estimator. appropriate an be can IPS distribution، target the to close is distribution proposal the
large a with means which unbiased، asymptotically is it but unbiased، not is estimator NIPS The
of number small a in Even value. target the to closer and closer becomes it samples، of number
as disadvantage، a necessarily not is NIPS in bias so values، target from far be may IPS samples،
has it that is NIPS of feature Another increases. samples of number the as diminishes it as long
of variance the ،q and p of choice the of independent So variance. its on bound upper fixed a
we but large is samples of number the when So property. favorable a is which small، is NIPS
estimator NIPS the ،q and p of closeness the about information or assumptions any know don't
choice. good a be can

with different are q and p when bias on bound lower positive fixed a has estimator PM The
cor- be not can estimator PM the in bias the IPS، and NIPS to contrast in So ratio. big some

can it other، each to close not are q and p if So have. we samples many how matter no rected variance its on bound a has estimator PM the But accurately. value target the estimate never variance the decrease arbitrarily can we So increases. samples of number the as vanishes which ،q and p of choice the of independent samples، of number the increasing by estimator the of small a with estimator stable a for So have. estimators NIPS and IPS the of none property a its of composition a is estimator an of MSE the that Note choice. good a be can PM variance، variance. and bias

vari- and bias small both has NIPS bias، on focused fully is estimator IPS the that see we Here finding merely is estimation the of goal the When variance. on more focused is PM and ance، sometimes But estimators. proposed the between chosen be can NIPS target، the of value the of value loss the is value target the e.g. value، target a maximizing or minimizing is goal the so important، less is estimator the in bias situation، this In model. a training for objective the choice. better a be can estimator PM the

worst the be can sometimes estimator unbiased the choosing that see can we conclusion، In bias. just than important more is variance and bias of composition the and do to thing