frequentist فرض کنید پارامتری از جامعه که به دنبال تخمین مقدار آن هستیم یک مجهول غیر تصادفی است که از طریق نمونه گیری به تعداد زیاد قابل تخمین است. به عنوان مثال با این شیوه اگر بخواهیم میانگین یک جامعه را تخمین بزنیم، یک نمونه تصادفی از افراد آن جامعه را انتخاب می کنیم، جواب های آنها را یادداشت می کنیم و میانگین آنها را به عنوان میانگین جامعه انتخاب می کنیم.

بازه اطمینان و P-value

Bayesian فرض می کند پارامتر مورد نظر ما یک متغیر تصادفی است. بنابراین در این شیوه بحث درباره تخمین مقدار پارامتر نیست بلکه درباره تخمین توزیع آن است. در این شیوه ابتدا یک حدس اولیه درباره توزیع متغیر تصادفی مورد نظرمان می زنیم و سپس با آزمایش و مشاهده یک نمونه، توزیع جدید آن را با کمک قانون بیز بدست می آوریم. به عبارت دیگر اگر متغیر تصادفی مورد علاقه ما θ باشد و متغیر تصادفی مورد آزمایش X به θ بستگی داشته باشد آنگاه داریم

Conjugate Prior و MAP

$$\underbrace{P(\theta \mid X)}_{\text{نتیجه گیری جدید از آزمایش}} = \frac{P(X \mid \theta)\, \underbrace{P(\theta)}_{\text{حدس اولیه}}}{P(X)}$$

update می کنیم

آماره های معروف

آماره: (static) st هر تابعی که روی یک نمونه تصادفی اعمال شود.

آماره تابعی از تعدادی متغیر تصادفی است پس خود نیز یک متغیر تصادفی است.

میانگین نمونه: $$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

امید ریاضی: $$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu$$

واریانس به شرط استقلال نمونه ها: $$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n \sigma^2 = \frac{1}{n} \sigma^2$$

Biased Variance: $$\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

UnBiased Variance $$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

تمرین: گفتیم اگر نمونه‌گیری ما به صورت بدون جایگذاری از مجموعه متناهی باشد دیگر متغیرهای تصادفی ما مستقل نیستند. ثابت کنید در این حالت: $Var(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}$ همانطور که از این روابط مشخص است واریانس میانگین نمونه در صورتی که نمونه‌گیری به صورت بدون جایگذاری باشد با ضریب $\frac{N-n}{N-1}$ کاهش می‌یابد و در صورتی که با n ثابت N به بی‌نهایت میل کند این ضریب برابر با 1 شده و بین نمونه‌گیری بدون جایگذاری و با جایگذاری تفاوتی نخواهد بود. راهنمایی: $Cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$

In the above, the covariance term is obtained as follows, because each of the $N(N-1)$ possible outcomes for $(y_1 - \mu)(y_2 - \mu)$ is equally likely:

$$\begin{aligned}
\operatorname{cov}(y_1, y_2) &= E\left((y_1 - \mu)(y_2 - \mu)\right) \\
&= \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} (x_i - \mu)(x_j - \mu) \\
&= \frac{1}{N(N-1)} (-N\sigma^2) \\
&= -\frac{\sigma^2}{N-1}
\end{aligned}$$

If we assume the simple random sampling is **without replacement**, then the sample values are **not** independent, so the covariance between any two different sample values is **not** zero. In fact, one can show that

Covariance between two different sample values: $\operatorname{cov}(X_i, X_j) = -\dfrac{\sigma^2}{N-1}$ for $i \neq j$

This fact is used to derive these formulas for the standard deviation of the estimator and the estimated standard deviation of the estimator. The first two columns are the parameter and the statistic which is the unbiased estimator of that parameter.

		standard deviation of the estimator	estimator of the standard deviation of the estimator
μ	$\overline{X}$	$\sqrt{\dfrac{\sigma^2}{n}\left(1-\dfrac{n-1}{N-1}\right)}$	$\sqrt{\dfrac{s^2}{n}\left(1-\dfrac{n}{N}\right)}$ where $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
p	$\hat{p}$	$\sqrt{\left(\dfrac{p(1-p)}{n}\right)\left(1-\dfrac{n-1}{N-1}\right)}$	$\sqrt{\dfrac{\hat{p}\left(1-\hat{p}\right)}{n}\left(\dfrac{n}{n-1}\right)\left(1-\dfrac{n}{N}\right)}$

تخمین
- پارامتری
  - نقطه‌ای
  - بازه‌ای
- غیر پارامتری

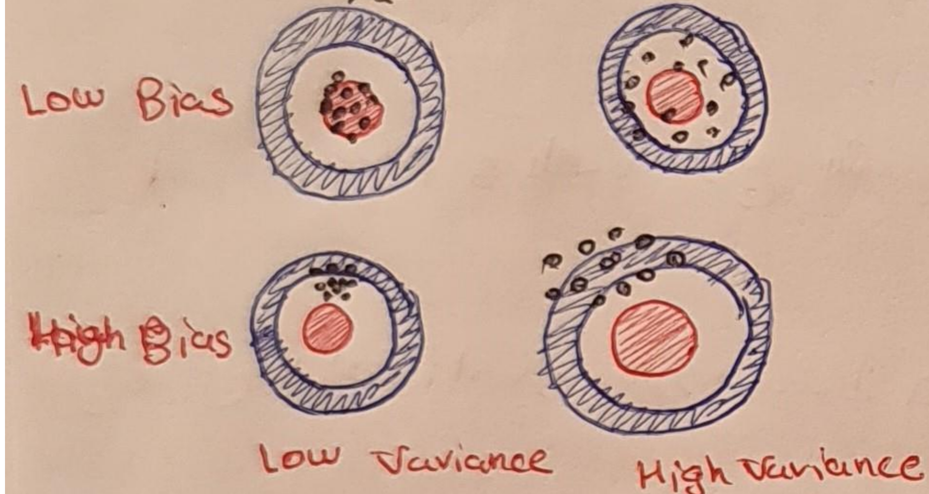
تخمین پارامتری: در این تخمین ما فرض می‌کنیم که مدلی که متغیر تصادفی ما از آن تبعیت می‌کند با یک بردار با بعد متناهی قابل توصیف است. مثلا $N(\mu,\sigma^2)$

← هر حالتی جز این غیر پارامتری است.

تخمین پارامتری نقطه‌ای ← $\hat{\theta}$ ، تخمین‌گری برای پارامتر θ باشد. به مقدار $E[(\hat{\theta}-E[\hat{\theta}])^2]$ ، $Var(\hat{\theta})$ می‌گوییم.

تعریف: فرض کنید که $\hat{\theta}$ تخمین‌گری برای پارامتر θ باشد. به مقدار $E[\hat{\theta}]-\theta$ ، $bias(\theta)$ می‌گوییم.

← به تخمین‌گری که bias آن برای تمامی مقادیر θ برابر با صفر باشد، unbiased می‌گوییم.



* واریانس در واقع متمرکز بودن و bias نزدیکی میانگین به هدف را می‌سنجد.

mean squared error

$$MSE(\hat{\theta}) = E[(\hat{\theta}-\theta)^2]$$

قضیه: $MSE(\hat{\theta}) = Var(\hat{\theta}) + bias^2(\hat{\theta})$

اثبات. مقدار θ ثابت است، $Var(\hat{\theta}) = Var(\hat{\theta}-\theta)$

از طرفی برای هر متغیر تصادفی Z ، $Var(Z)=E[Z^2]-(E(Z))^2$ حالا $Z=\hat{\theta}-\theta$

$$\Rightarrow Var(\hat{\theta}) = Var(\hat{\theta}-\theta) = \underbrace{E[(\hat{\theta}-\theta)^2]}_{MSE(\hat{\theta})} - (\underbrace{E(\hat{\theta}-\theta)}_{bias(\hat{\theta})})^2$$

تخمین‌گر سازگاری Consistency :

تخمین‌گرهایی که ما می‌سازیم، تابعی از n که تعداد داده‌ها است، می‌باشند و ما انتظار داریم که با افزایش تعداد داده‌ها، شناختمان نسبت به توزیع بهتر شود.

* $\hat{\theta}$ سازگار است هرگاه

$$\text{for all } \varepsilon>0 ,\ \lim_{n\to\infty} P(|\hat{\theta}_n-\theta|\geq\varepsilon)=0$$

* در مثال پواسون، تخمین‌گر $T(x_1,\ldots,x_n)=\frac{\sum_{i=1}^{n} x_i}{n}$ سازگار است ولی تخمین‌گر $T(x_1,\ldots,x_n)=x_n$ سازگار نیست.

$T(x_1,\ldots,x_n)=x_1$ یک تخمین‌گر می‌تواند unbiased باشد ولی ناسازگار باشد

$T(x_1,\ldots,x_n)=\frac{\sum x_i}{n}+\frac{1}{n}$ " " " biased " " سازگار "

مثال. فرض کنید n داده تصادفی به صورت $X_1,\ldots,X_n$ با میانگین $E[X_i]=\theta$ و واریانس $Var(X_i)=\sigma^2$ داریم. می‌خواهیم نشان دهیم $\hat{\theta}_n=\overline{X}$ یک تخمین‌گر سازگار θ است.

حکم: $\text{for all } \varepsilon>0 \quad \lim_{n\to\infty} P(|\hat{\theta}_n-\theta|\geq\varepsilon)=0$

چبیشف: $P(|\overline{X}-\theta|\geq\varepsilon)\leq\frac{Var(\overline{X})}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$ ، $n\to\infty \Rightarrow \frac{\sigma^2}{n\varepsilon^2}\to 0$

$\boxed{* \ Var(\overline{X})=\frac{\sigma^2}{n}}$

$$MSE(\hat{\theta})=E[(\hat{\theta}-\theta)^2]=E[(\overline{X}-\theta)^2]=\underbrace{Var(\overline{X}-\theta)}_{Var(\overline{X})}+\underbrace{E([\overline{X}-\theta])^2}_{0}$$

$$\Rightarrow MSE(\hat{\theta})=Var(\overline{X})=\frac{\sigma^2}{n} ,\ n\to\infty ,\ \frac{\sigma^2}{n}=0$$

قضیه. اگر ... $\hat{\theta}_1$ ، $\hat{\theta}_2$ ، سری از تخمین‌گرهای نقطه‌ای θ باشد، $\hat{\theta}_n$ را یک تخمین‌گر سازگار می‌نامیم اگر $\lim_{n\to\infty} MSE(\hat{\theta}_n)=0$

$$P(|\hat{\theta}_n-\theta|\geq\varepsilon)=P(|\hat{\theta}_n-\theta|^2\geq\varepsilon^2)\leq\underbrace{\frac{E[\hat{\theta}_n-\theta]^2}{\varepsilon^2}}_{0} \ \text{مارکوف} \ =\frac{MSE(\hat{\theta}_n)}{\varepsilon^2} \quad n\to\infty\Rightarrow$$

n-1 بجای n

n متغیر تصادفی به صورت $x_1, x_2, \_\_\_\_$ داریم. همانطور که می‌دانیم اگر امید ریاضی آنها μ باشد برای محاسبه‌ی واریانس از رابطه‌ی زیر استفاده می‌کنیم $\sigma^2 = E[(X-\mu)^2]$ اما ما می‌خواهیم μ را بیابیم ← پس به جای μ از $\bar{X}$ استفاده می‌کنیم.

$$\bar{S}_1^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X})^2 \longrightarrow \text{bias}$$ دارد :

$$E(\bar{X}^2) = E(\bar{X})^2 + Var(\bar{X}) = \frac{\sigma^2}{n} + \mu^2$$

$$E(X_i^2) = E(X_i)^2 + Var(X_i) = \sigma^2 + \mu^2$$

$$E(\bar{S}_1^2) = \frac{1}{n}\left(\sum_{k=1}^{n}E(X_i^2) - nE(\bar{X}^2)\right)$$

$$= \frac{1}{n}\left(n(\mu^2+\sigma^2) - n\left(\mu^2+\frac{\sigma^2}{n}\right)\right) = \frac{n-1}{n}\sigma^2$$

* $$\text{Bias}(\bar{S}_1^2) = E(\bar{S}_1^2) - \sigma^2 = -\frac{\sigma^2}{n}$$

* n خوب نبود یک estimator ⟸

بجای : $$\bar{S}_2^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2$$

liklihood

MLE   Maximum Liklihood Estimator تخمین‌گر بیشینه

هدف ما تخمین پارامتر θ است. برای این کار، برای مقادیر مختلف θ، احتمال داده ها رو حساب می‌کنیم و بر آن تابع درست‌نمایی یا می‌گوییم.

* گسسته : $\mathcal{L}(\theta|x) = P_\theta(x)$ * پیوسته : $\mathcal{L}(\theta|x) = f_\theta(x)$

* تخمین‌گر درست‌نمایی بیشینه، تخمین‌گری است که با مشاهده داده، مقدار θ ای که برای آن تابع $\mathcal{L}(\theta|x)$ بیشینه است را برای تخمین خود از θ معرفی می‌کند.

لگاریتم درست‌نمایی

بسیاری از مواقع، ما از یک توزیع تعداد زیادی داده مستقل داریم. برای یافتن تخمین‌گر بیشینه باید عبارت $\prod P(x_i|\theta) = \mathcal{L}(\theta|D)$ را بیشینه کنیم.
اکثر مواقع برای یافتن θای که عبارت فوق را بیشینه کند، باید از آن مشتق گرفت که وقت‌گیر است به خاطر همین کافی است از $\ln$ آن مشتق بگیریم.

* تخمین MLE برای $\hat{\mu}_{ML} = \frac{\sum_{i=1}^{N} x^{(i)}}{N}$ و $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x^{(i)} - \hat{\mu}_{ML})^2$

مثال. فرض کنید که داده های x_n ، ... ، x_1, x_2 ، n نمونه مستقل از توزیع نرمال $(\theta, 1)$ می باشد. تخمین گر درست نمایی بیشینه θ را بیابید.

$$\sum \frac{d}{d\theta}\left(\ln P(x_i|\theta)\right) = 0 \Rightarrow P(x_i|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} \longrightarrow \ln(\sqrt{2\pi}) - \frac{(x_i-\theta)^2}{2}$$

$$\longrightarrow \frac{d}{d\theta} \ln P(x_i|\theta) = x_i - \theta \Rightarrow \sum x_i - n\theta_{ML} = 0 \longrightarrow \boxed{\theta_{ML} = \frac{\sum x_i}{n}}$$

* $\mathcal{L}(\theta_1, \theta_2 | x_1, \ldots x_n) = P(x_1 \ldots x_n ; \theta_1, \theta_2) = P(x_1 ; \theta_1, \theta_2) P(x_2 ; \theta_1, \theta_2) \ldots P(x_n ; \theta_1, \theta_2)$

چون i.i.d

* $\underset{\theta_1, \theta_2}{\arg\max} (f(\theta_1, \theta_2)) \longrightarrow$ چه θ_1 و θ_2 ای f رو max می کنه *

* ویژگی های تخمین گر بیشینه MLE

* اگر $\hat{\theta}$ تخمین گر بیشینه باشد:

* unbiased $\lim_{n\to\infty} \mathbb{E}(\hat{\theta}) = \theta$

* Consistent $\forall_{\varepsilon>0}: \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) \lessgtr 0$

* با بیشتر شدن مقدار n ، $\hat{\theta}$ تبدیل به یک متغیر تصادفی نرمال می شود.

$$\frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} \sim N(0, 1)$$

گام اول برای بازه اطمینان

* Frequentist ← MLE

baysian ← MAP

Maximum a Posteriori

چارک های نرمال استاندارد

متغیر تصادفی نرمال استاندارد را Z، تابع توزیع تجمعی آن را Φ بنامید. Z_α را به ازای $0<\alpha<1$ تعریف می کنیم.

$P(Z>Z_\alpha)=\alpha$ که معادل است با $P(Z\leqslant Z_\alpha)=1-\alpha$ بنابر این داریم $Z_\alpha=\Phi^{-1}(1-\alpha)$ ← این مقدار چارک نرمال استاندارد نام دارد

$qnorm(1-\alpha)=Z_\alpha$

$*\ P(Z<-Z_\alpha)=P(Z>Z_\alpha)=\alpha$, $P(|Z|>Z_\alpha)=2\alpha \Rightarrow P(-Z_\alpha\leqslant Z\leqslant Z_\alpha)=1-2\alpha$

$*\ 1-\alpha=P(-Z_{\alpha/2}\leqslant Z\leqslant Z_{\alpha/2})$

بازه اطمینان

می دانیم که تخمین بیشینه درست نمایی برای پارامتر میانگین μ برابر است با میانگین نمونه ای $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$، اما این تخمین چقدر دقیق است؟

اگر داده های ما به صورت مستقل و یکسان توزیع شده باشد (iid) از توزیع نرمال بدست آمده باشند:

$$\hat{\mu}_{MLE}=\frac{x_1+\cdots+x_n}{n}\sim N\left(\mu,\frac{\sigma^2}{n}\right)\xrightarrow[\hat{\mu}\ \text{استاندارد کردن}]{}\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

$$\Rightarrow P\left(-Z_{\alpha/2}\leqslant\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}\leqslant Z_{\alpha/2}\right)=1-\alpha.$$

$$\Rightarrow P\left(\hat{\mu}-\left(Z_{\alpha/2}\times\frac{\sigma}{\sqrt{n}}\right)\leqslant\mu\leqslant\hat{\mu}+\left(Z_{\alpha/2}\times\frac{\sigma}{\sqrt{n}}\right)\right)=1-\alpha$$

* یعنی احتمال اینکه پارامتر μ در بازه ی بالا باشد $1-\alpha$ است.

بازه اطمینان

مثال. فرض کنید می خواهیم بازه ی اطمینان 95% ی را برای تخمین از پارامتر μ در مثال بالا بدست آوریم.

با R می دانیم $Z_{0.025}=1.96$ $\Rightarrow P\left(-1.96\leqslant\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}\leqslant 1.96\right)=0.95$

بنابراین بازه ی اطمینان 95 درصدی برای پارامتر μ برابر است با $\left[\hat{\mu}-1.96\frac{\sigma}{\sqrt{n}},\hat{\mu}+1.96\frac{\sigma}{\sqrt{n}}\right]$

فرض کنید توزیع $uniform(a, b)$ داده شده است.

الف

۳ داده ی ۱۰ و ۸ و ۱۲/۵ از این توزیع انتخاب شده اند. maximum likelyhood را برای پارامترهای a و b تخمین بزنید

ب

اگر n داده از این توزیع به صورت $x_۱, x_۲, ..., x_n$ انتخاب شده باشند، maximum likelyhood را برای پارامترهای a و b تخمین بزنید

فرض کنید توزیع uniform(a, b) داده شده است.

الف) 3 داده ی 10 و 8 ، 12,5 از این توزیع انتخاب شده اند. MLE را برای پارامترهای a ، b تعیین کنید.

ب) اگر n داده از این توزیع به صورت $x_1, x_2, \ldots, x_n$ انتخاب شده باشند، MLE را برای پارامترهای a ، b بدست آورید.

$$MLE(a,b) = \left(\frac{1}{b-a}\right)^n \quad , \quad a \leq x_1, \ldots, x_n \leq b$$

همان طور که می دانیم باید ماکسیمم شود

پس b-a باید کمترین مقدار خود شود

$$\Rightarrow a_{MLE} = \min(x_1, x_2, \ldots, x_n)$$

$$b_{MLE} = \max(x_1, x_2, \ldots, x_n)$$

برای الف هم

$$\begin{cases} a = 8 \\ b = 12.5 \end{cases}$$

مسئله‌ی ۳. تاس های نامتقارن

یک تاس k وجهی به شما داده است. همچنین یک مجموعه نمونه از نتایج پرتاب این تاس در اختیار داریم. تعداد آمدن هر وجه تاس در مجموعه m_k است. به صورتی که

$$\sum_{i=۱}^{i=k} m_i = N$$

احتمال آمدن هر وجه تاس نیز با μ_i نمایش داده می‌شود. همچنین $X(x) = k$ نشان دهنده مقدار متغیر تصادفی وجه تاس است.(به عنوان مثال $X(\omega) = ۱$ یعنی وجه اول آمده است) تعداد کل پرتاب ها انجام شده N تعداد می‌باشد.

الف

فرض کنید که تاس دو وجهی است یک تخمین گر بیشینه درست نمایی برای احتمال هر وجه پیدا کنید.(دقت کنید که راه حل شما باید حتما شامل maximum likelihood باشد و به اینکه از نظر فیزیکی آیا تاس ۲ وجهی ممکن است یا خیر نیز توجه نداشته باشید.)

مسئله 7. تاس های نامتقارن

یک تاس K وجهی به شما داده شده است؛ هم چنین یک مجموعه نمونه از نتایج پرتاب این تاس در اختیار داریم. تعداد باری که وجه i ام تاس ظاهر شده، m_i است و می‌دانیم $\sum_{i=1}^{K} m_i = N$. احتمال آمدن وجه i ام تاس نیز با μ_i نمایش داده می شود و همچنین $X(\omega)=i$ نشان دهنده مقدار متغیر تصادفی وجه رو آمده از تاس است (به عنوان مثال $X(\omega)=1$ یعنی وجه اول تاس آمده است). تعداد کل پرتاب های انجام شده، N تا است.

الف) فرض کنید که تاس، دو وجهی است؛ یک تخمین گر بیشینه ی درست نمایی برای احتمال رو آمدن هر وجه پیدا کنید.

$$D = \{ x^{(1)}, x^{(2)}, \_\_\_, x^{(N)} \} \quad \text{و} \quad m_1 \text{ تا } X(1) \quad \text{و} \quad m_2 \text{ تا } X(2)$$

$$P(x|\theta) = \theta^{x}(1-\theta)^{1-x}$$

$$\boxed{\mu_1 = \theta} \longrightarrow \text{bernouli}$$

$$\Rightarrow P(D|\theta) = \prod_{i=1}^{N} p(x^{(i)}|\theta) = \prod_{i=1}^{N} \theta^{x^{(i)}}(1-\theta)^{1-x^{(i)}} \Rightarrow \ln P(D|\theta) = \sum_{i=1}^{N} \{ x^{(i)} \ln\theta + (1-x^{(i)}) \ln(1-\theta) \}$$

$$\frac{\partial \ln P(D|\theta)}{\partial\theta} = 0 \Rightarrow \frac{\partial(\ln\theta \; m_1 + (N-m_1)\ln(1-\theta))}{\partial\theta} = \frac{m_1}{\theta} + \frac{N-m_1}{\theta-1} = \frac{m_1\theta - m_1 + N\theta - m_1\theta}{\theta(\theta-1)} = 0$$

$$\Rightarrow N\theta = m_1 \Rightarrow \boxed{\theta = \frac{m_1}{N}} \Rightarrow \mu_1 = \frac{m_1}{N}, \quad \mu_2 = \frac{m_2}{N}$$

ب) این بار، فرض کنید که K وجه داریم و مسئله را بصورت کلی حل کنید، یعنی یک تخمین گر بیشینه درست نمایی برای هر یک از μ_K بدست آورید.

multinomial distribution , Parameter space : $\mu = [\mu_1, \_\_\_, \mu_K]$

$$\sum_{i=1}^{K} \mu_i = 1$$

$$P(x|\cancel{\theta}\mu) = \prod_{k=1}^{K} \mu_k^{x_k}$$

$$D = \{x^{(1)}, x^{(2)}, \_\_\_, x^{(N)}\} \Rightarrow P(D|\mu) = \prod_{i=1}^{N} P(x^{(i)}|\mu) = \prod_{i=1}^{N}\prod_{k=1}^{K} \mu_k^{x_k^{(i)}} = \prod_{k=1}^{K} \mu_k^{m_k}$$

$$\sum_{i=1}^{K} \mu_i = 1 \quad \text{و} \quad \sum_{k=1}^{K} m_k = N \Rightarrow$$

Lagrange multiplier : $\sum_{k=1}^{K} m_k \ln \mu_k + \lambda\left(1 - \sum_{k=1}^{K} \mu_k\right) = L(\mu, \lambda)$

$$\frac{\partial L(\mu, L)}{\partial \mu_i} = 0 \Rightarrow \boxed{\mu_i = -\frac{m_i}{\lambda}}$$

و چون $\sum \mu_i = 1 \Rightarrow \boxed{\lambda = -\sum m_i = -N}$

$$\Rightarrow \hat{\mu}_i = \frac{m_i}{N}$$

مسئله 6 و آمار دان

* داده های $x^{(n)}, \dots, x^{(2)}, x^{(1)}$ نمونه هایی از متغیر تصادفی گسسته با مقادیر طبیعی 1 تا k هستند، برای بردار احتمال $\vec{P} = (P_1, \longrightarrow P_k)$ تخمین گر درستی بیشینه را محاسبه کنید. فرض کنید که مقادیر 1 تا k، هرکدام α_i بار مشاهده شوند در اینصورت

$$L(P|D) = \prod p_i^{\alpha_i}$$

بنابراین باید

$$\prod p_i^{\alpha_i} = \underbrace{\prod \left(\frac{p_i}{\alpha_i}\right)^{\alpha_i}}_{} \cdot \underbrace{\prod \alpha_i^{\alpha_i}}_{\text{ثابت}} \quad , \sum \alpha_i = n$$

$$\xrightarrow{\text{AM-GM}} \leq \frac{\sum \alpha_i \frac{p_i}{\alpha_i}}{n} = 1$$

، تساوی وقتی رخ می دهد که همه برابر باشند $\longrightarrow \hat{P}_{MLE} = \left(\frac{\alpha_1}{n}, \frac{\alpha_2}{n}, \dots, \frac{\alpha_k}{n}\right)$

مسئله‌ی ۴. بازه‌های شکست *

یک متغیر برنولی با احتمال موفقیت p را در نظر بگیرید که از مقدار p مطلع نیستیم. حال متغیر T را به صورت زیر برای هر موفقیت تعریف می‌کنیم:
T_i برابر تعداد شکست های بین موفقیت $T_{i-۱}$ و T_i، با احتساب موفقیت K ام است؛ به عبارت دیگر،

$$T_1 = Y_1 \, , T_k = Y_k - Y_{k-1}, \; k = 1, 2, 3, ...$$

که متغیر Y_K برابر شماره‌ی آزمایشی است که K امین موفقیت را در پی داشته است. تعداد کل آزمایش‌ها به ما داده نشده است، بلکه تنها مجموعه‌ی T را در اختیار داریم که برابر $T = \{T_۱, T_۲, ..., T_k\}$ است.

الف) یک تخمین‌گر بیشینه‌ی درست‌نمایی برای پارامتر P به دست آورید. (دقت کنید که راه‌حل شما باید به صورت maximum likelihood باشد)

ب) با فرض دانستن p به نظر شما توزیع T_i به چه توزیع‌ای شبیه است ($PMF_{T|p}$)؟ میانگین و واریانس این توزیع را به دست آورید. سپس با استفاده از توزیع T و با فرض ندانستن p و هم‌چنین بدون هیچ پیش‌فرضی از داده‌های به دست آمده (یعنی فرض کنید که هیچ داده‌ای نیز به شما داده نشده است)، تابع PMF_T را به دست آورید.

ج) نشان دهید که رابطه‌ی زیر برقرار است (p^* برابر مقدار واقعی پارامتر توزیع است):

$$\forall \epsilon > ۰ \; : \; {}_{k \to \infty} P(|\tfrac{1}{P} - \tfrac{1}{p^*}| > \epsilon) = ۰$$

د) فرض کنید که $p^* > ۰/۵$ است. به کمک نامساوی چبی‌شف، حال یک کران پایین برای K بیابید به صورتی که رابطه‌ی زیر برقرار باشد:

$$P(|\tfrac{1}{P} - \tfrac{1}{p^*}| \leqslant ۰/۱) \geqslant ۰/۹۵$$

مسئله 4. بازه های شکست

یک متغیر برنولی با احتمال موفقیت p را در نظر بگیرید که از مقدار p مطلع نیستیم. حال متغیر T را بصورت زیر برای هر موفقیت تعریف می‌کنیم:
T_i برابر تعداد شکست های بین موفقیت T_{i-1} ، T_i ، با احتساب موفقیت k ام است؛ به عبارت دیگر،

$$T_1 = Y_1 \quad , \quad T_k = Y_k - Y_{k-1} \quad , \quad k = 1, 2, 3, \ldots$$

که متغیر Y_k برابر شماره‌ی آزمایشی است که k امین موفقیت را در پی داشته است.
بلکه تنها مجموعه T را در اختیار داریم که برابر $T = \{T_1, T_2, \ldots, T_k\}$ است. تعداد کل آزمایش ها به ما داده نشده است.

الف) یک تخمین گر بیشینه ی درست نمایی برای پارامتر p بدست آورید.

$$T_i = (1-p)^{t_i - 1} p \Rightarrow P\{T_1, T_2, \ldots, T_k \mid p\} = \arg\max_p (1-p)^{|\sum t_i - k|} p^k$$

$$= \ln L(T; p) = |\sum t_i - k| \ln(1-p) + k \ln(p)$$

$$\frac{\partial L}{\partial p} = \frac{\sum t_i - k}{p-1} + \frac{k}{p} = \hat{p}_{ML} = \frac{k}{\sum t_i}$$

ب) با فرض دانستن p به نظر شما توزیع T_i به چه توزیعی شبیه است $(PMF_{T|p})$؟ میانگین و واریانس این توزیع را بدست آورید. سپس با استفاده از توزیع T و با فرض ندانستن p و همچنین بدون هیچ پیش فرضی از داده های بدست آمده (یعنی فرض کنید که هیچ داده ای نیز به شما داده نشده) تابع PMF T را بدست آورید.

توزیع ~~[illegible]~~ هندسی و $\mu = \frac{1}{p}$ ، $Var = \frac{1-p}{p^2} \Leftarrow$

$$PMF_{T|p}(t) = (1-p)^{t-1} p$$
$$t = 1, 2, \ldots$$

$$PMF_T(t) = \int_0^1 PMF_{T|p} \, dp = \int_0^1 (1-p)^{t-1} p \, dp = \left[-\frac{(1-p)^t (tp+1)}{t(t+1)} \right|_0^1 = \frac{1}{t(t+1)}$$

ج) نشان دهید که رابطه زیر برقرار است. (p^* برابر مقدار واقعی پارامتر توزیع است)؛ $\forall \varepsilon > 0 : k \to \infty \; P(|\frac{1}{p} - \frac{1}{p^*}| > \varepsilon) = 0$

از آنجا که مقدار $\frac{1}{p}$ برابر مقدار بدست‌آمده برای میانگین داده هاست و مطابق قانون اعداد بزرگ (دقت کنید که همه T_1 یک‌اند.) مقدار میانگین یک متغیر تصادفی با میانگین بدست‌آمده از داده های پایدار است و بایاس ندارد به عبارتی

$$\forall \varepsilon > 0 : \lim_{k \to \infty} P\left(\left|\frac{\sum_{t_i \in T} t_i}{k} - E(T_1)\right| > \varepsilon\right) = 0$$

د) فرض کنید که $p^* > 0.5$ است. به کمک نامساوی چبی‌شف، حال یک کران پایین برای k بیابید که $P(|\frac{1}{p} - \frac{1}{p^*}| \leq 0.1) \geq 0.95$

$$P\left(\left|\frac{\sum_{t_i \in T} t_i}{k} - E[T_1]\right| > \varepsilon\right) \leq \frac{var(T_1)}{k\varepsilon^2}$$

$$P\left(\left|\frac{\sum_{t_i \in T} t_i}{k} - \frac{1}{p^*}\right| \leq 0.1\right) = 1 - P\left(\left|\frac{\sum_{t_i \in T} t_i}{k} - \frac{1}{p^*}\right| > 0.01\right) \geq 1 - \frac{var(T_1)}{k \times 0.01}$$

$$1 - \frac{var(T_1)}{k \times 0.01} \geq 0.95 \Rightarrow k > 2000\left(\frac{1-p}{p^2}\right)$$

حداکثر عبارت $\frac{1-p}{p^2}$ برابر 2 است. چون p بین $1, \frac{1}{2}$ است $\Leftarrow k \geq 4000$

در چرنوبیل یک منبع در هر زمانی که اندازه گیری می شود K فوتون از خود ساطع می کند . ما فرض می کنیم که K دارای توزیع زیر است :

$$p_K(k;\theta) = c(\theta)e^{-k\theta} \qquad k = ۰, ۱, ۲, ...$$

بطوری که θ معکوس دمای منبع است و $c(\theta)$ ضریب نرمالیزاسیون است . همچنین فرض میکنیم که فوتون های ساطع شده از منبع هر بار مستقل از یکدیگر هستند . ما می خواهیم دمای منبع را با اندازه گیری های پی در پی فوتون های ساطع شده تخمین بزنیم .

الف

ضریب نرمالیزاسیون $c(\theta)$ را بیابید.

ب

تخمین ML را برای دمای منبع $\psi = \frac{۱}{\theta}$ را بر اساس $K_۱, K_۲, ...$ که تعداد فوتون های ساطع شده پس از n بار اندازه گیری است را بیابید.

مسئله 3. چرنوبیل

منبعی رادیواکتیو در چرنوبیل وجود دارد که در هر زمانی که اندازه گیری شود k، فوتون از خود ساطع می کند. ما فرض می کنیم که K دارای توزیع زیر است:

$$P_K(k;\theta) = C(\theta)\, e^{-k\theta} \qquad k=0,1,2,\ldots$$

به طوری که θ معکوس دمای منبع بوده و $C(\theta)$ کمیتی به نام ضریب نرمالیزاسیون است. هم چنین فرض می کنیم که فوتون های ساطع شده از منبع هربار مستقل از یکدیگراند. ما می خواهیم دمای منبع را با اندازه گیری های پی در پی فوتون های ساطع شده، تخمین بزنیم.

الف) ضریب نرمالیزاسیون $C(\theta)$ را بیابید.

اگر θ را مقداری مشخص در نظر بگیریم طبق اصول کولموگروف

$$\sum_{i=0}^{\infty} P_K(i;\theta) = 1 \Rightarrow \sum_{i=0}^{\infty} c(\theta)\, e^{-i\theta} = c(\theta)\sum_{i=0}^{\infty} e^{-i\theta} = c(\theta)\sum_{i=0}^{\infty} x^i$$

$$e^{-\theta} = x$$

$$= \frac{x^{\infty}-1}{x-1}\times c(\theta) = \frac{0-1}{e^{-\theta}-1}\times c(\theta) = 1 \Rightarrow c(\theta) = 1-e^{-\theta}$$

ب) تخمین ML برای دمای منبع $T=\frac{1}{\theta}$ را بر اساس $k_1,k_2,\ldots$ (که k_i تعداد فوتون های ساطع شده در اندازه گیری iاُم است) را بیابید.

$$ML(k_1,k_2,\ldots,k_n|\theta) = \underbrace{\prod_{i=1}^{n}}_{\text{چون مستقل اند}} P(k_i|\theta)$$

می خواهیم بیشینه شود

$$\Rightarrow \hat{\theta}_{ML} = \arg\max_{\theta} \sum_{i=1}^{n} \frac{d}{d\theta}\ln P(k_i|\theta) = 0 \Rightarrow \ln P(k_i|\theta) = \ln(1-e^{-\theta}) + (-k_i\theta)$$

$$\Rightarrow \frac{d}{d\theta} = \frac{1}{e^{\theta}-1} - k_i$$

$$\Rightarrow \frac{n}{e^{\theta}-1} - \sum k_i = 0 \Rightarrow \frac{1}{e^{\theta}-1} = \frac{\sum k_i}{n} \Rightarrow mean(k_1,\ldots,k_n) = \bar{k}$$

$$\Rightarrow e^{\theta}-1 = \frac{1}{\bar{k}} \Rightarrow \theta = \ln\left(\frac{1}{\bar{k}}+1\right)$$

$$\Rightarrow T = \frac{1}{\ln\left(\frac{1}{\bar{k}}+1\right)}$$

توزیع Pareto با دو پارامتر α و x_m مشخص می‌شود. می‌دانیم که $x_۱$، $x_۲$، ... و x_n داده‌هایی رندوم از توزیع Pareto با $\alpha > ۲$ هستند. خصوصیات توزیع Pareto داده‌هایمان در ادامه آورده شده است:

$$PDF: \ \frac{\alpha {x_m}^{\alpha}}{x^{\alpha+۱}}, \ \alpha > ۲, \ x_m > ۰, \ x \geqslant x_m$$

$$CDF: \ ۱ - (\frac{x_m}{x})^{\alpha}$$

$$Mean: \ \frac{\alpha x_m}{\alpha - ۱}$$

$$Variance: \ \frac{\alpha {x_m}^{۲}}{(\alpha - ۱)^{۲}(\alpha - ۲)}$$

الف

با استفاده از تخمین‌گر MLE پارامترهای α و x_m را تخمین بزنید.

ب

وضعیت unbiased بودن و consistent بودن تخمین‌گر MLE برای x_m را مشخص کنید.
(راهنمایی: اگر n متغیر تصادفی iid از توزیعی با $CDF: \ F$ باشند، مینیموم این متغیرها متغیر تصادفی‌ای با $CDF: \ ۱ - (۱ - F)^n$ است)

ابتدا x_m را تخمین می‌زنیم. باید عبارت زیر را بیشینه کنیم:

$$\sum_{i=1}^{n} ln(P(x_i|x_m)) = \sum_{i=1}^{n} ln(\frac{\alpha {x_m}^{\alpha}}{{x_i}^{\alpha+1}}) = nln(\alpha) + n\alpha ln(x_m) - (\alpha + 1)(\sum_{i=1}^{n} ln(x_i))$$

برای بیشینه کردن این عبارت بر حسب x_m کافی است x_m را بیشینه کنیم. برای هر i می‌دانیم $x_i \geqslant x_m$. پس $min(x_۱, ..., x_n) \geqslant x_m$. پس تخمین‌مان برای x_m برابر $min(x_۱, ..., x_n)$ است.

حال α را تخمین می‌زنیم. باید عبارت زیر را بیشینه کنیم:

$$\sum_{i=۱}^{n} ln(P(x_i|\alpha)) = nln(\alpha) + n\alpha ln(x_m) - (\alpha + ۱)(\sum_{i=۱}^{n} ln(x_i))$$

مشتق این عبارت بر حسب α را برابر ۰ قرار می‌دهیم. در این صورت داریم:

$$\alpha = \frac{n}{\sum_{i=۱}^{n} ln(x_i) - nln(x_m)}$$

ب

CDF تخمین‌گر $min(x_۱, ..., x_n)$ طبق راهنمایی برابر است با $۱ - (\frac{x_m}{x})^{\alpha n}$. همان‌طور که می‌بینید این عبارت CDF یک توزیع Pareto با پارامترهای $(\alpha n, x_m)$ است. پس MSE این تخمین‌گر برابر است با:

$$(\frac{\alpha n x_m}{\alpha n - ۱} - x_m)^۲ + \frac{\alpha n {x_m}^۲}{(\alpha n - ۱)^۲(\alpha n - ۲)} =$$

$$(\frac{x_m}{\alpha n - ۱})^۲ + \frac{\alpha n {x_m}^۲}{(\alpha n - ۱)^۲(\alpha n - ۲)}$$

حال اگر n را به بی‌نهایت میل دهیم، MSE به ۰ میل می‌کند. پس این تخمین‌گر consistent است.

اما چون داریم $bias = \frac{x_m}{\alpha n - ۱}$ پس unbiased است.

$\triangleright$

صورت کلی مسئله فضای احتمال $(\Omega, \mathcal{F}, \mathbb{P})$ را در نظر بگیرید. از این فضای احتمال، n نمونه تصادفی مستقل برداشته و آنها را به صورت $\omega_۱, \omega_۲, , \ldots, \omega_n$ نشان می‌دهیم.

الف

برای یک پیشامد A تخمین زیر را برای $\mathbb{P}(A)$ در نظر می‌گیریم.

$$\widehat{\mathbb{P}}(A) = \frac{\sum_{i=۱}^{n} I_A(\omega_i)}{n}$$

که در آن $I_A(\omega_i)$ یک است اگر ω_i در A باشد و در غیر این صورت صفر است. نشان دهید که این تخمین یک تخمین unbiased و سازگار برای $\mathbb{P}(A)$ می‌باشد. (۴ نمره)

ب

کران بالای مناسبی برای $\mathbb{P}(|\widehat{\mathbb{P}}(A) - \mathbb{P}(A)| > \epsilon)$ بنویسید.(هر کران بالای غیر بدیهی قابل قبول است.) (۲ نمره)

ابتدا ثابت می‌کنیم تخمینگر unbiased است.

$$\mathbb{E}[\widehat{\mathbb{P}}(A)] = \mathbb{E}[\frac{\sum_{i=1}^{n} I_A(\omega_i)}{n}] = \frac{1}{n}\mathbb{E}[\sum_{i=1}^{n} I_A(\omega_i)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[I_A(\omega_i)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(\omega_i \in A) = \mathbb{P}(A)$$

حالا نشان می‌دهیم که سازگار است. از قضیه نوت بوک‌ها در حالت unbiased کمک می‌گیریم.

$$MSE = Var(\widehat{\mathbb{P}}(A)) = Var(\frac{1}{n}\sum_{i=1}^{n} I_A(\omega_i)) = \frac{1}{n^2}\sum_{i=1}^{n} Var(I_A(\omega_i))$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{P}(\omega_i \in A) - \mathbb{P}(\omega_i \in A)^2 = \frac{\mathbb{P}(A)-\mathbb{P}(A)^2}{n}$$

لذا وقتی $n \to \infty$ داریم $MSE \to \infty$ پس سازگار است.

ب

از نامساوی چبیشف کمک می‌گیریم.

$$\mathbb{P}(|\widehat{\mathbb{P}}(A) - \mathbb{P}(A)| > \epsilon) \leqslant \frac{Var(\widehat{\mathbb{P}}(A))}{\epsilon^2} = \frac{\mathbb{P}(A)-\mathbb{P}(A)^2}{n\epsilon^2}$$

مسئله‌ی ۱. قطعه‌سازی

در کارخانه تولید اتومبیل، قطعه‌ای با وزن تصادفی تولید می‌شود. n نمونه مستقل از قطعه‌ها را انتخاب کرده‌ایم و وزن آن‌ها را $w^{(1)}, w^{(2)}, ..., w^{(n)}$ اندازه گرفته‌ایم. اگر بدانیم وزن قطعه از توزیع زیر به دست می‌آید، تخمینگر بیشینه درست نمایی برای پارامتر λ را به دست آورید. (۴ نمره)

$$f_X(x) = \lambda x e^{\frac{-\lambda x^2}{2}} U(x)$$

حل. توضیح کلی راه حل:

$$\ell_n(\lambda) = n \log \lambda + \sum_{i=1}^{n} \log w_i - \frac{\lambda}{2} \sum_{i=1}^{n} w_i^2 \Rightarrow \widehat{\lambda}_{ML} = \frac{2n}{\sum_{i=1}^{n} w_i^2}$$