جلسه‌ی بیستم آمار و احتمال مهندسی دکتر شریفی زارچی دانشگاه شریف

در جلسه‌ی قبل متوجه هدف علم آمار شدیم که در تلاش بود از جزء به کل برسه. حالا ما دنبال پارامترهایی مثل μ، σ^2 در کل جامعه هستیم. دست به تخمین زدن می‌زنیم. این جلسه کلاً به روش‌های پیدا کردن μ و σ^2 می‌پردازیم.

پارامترهای توزیع (تخمین)

به دنبال پیدا کردن μ، σ^2 در کل جامعه هستیم در حالتی که توزیع داده‌ها رو می‌دونیم. یا توزیع رو حدس زدیم و یا از روش‌های دیگه بدست آوردیم و با توجه به داده‌های sample به دنبال تخمین پارامترهای توزیع در کل جامعه هستیم.

تخمین نقطه‌ای: پارامتر رو از یک راه حل مثل میانگین یا واریانس یا ... از داده‌های sample بدست می‌آوریم.

تخمین بازه‌ای: همون تخمین نقطه‌ای رو در یک بازه مثلا ± 1 گزارش می‌کنیم و می‌گیم با احتمال مثلاً α درصد، پارامتر ما داخل این بازه است. هدف ما پیدا کردن کوچک‌ترین بازه با بیشترین احتمال است.

پارامتر θ از جمعیت که نمی‌دونیم. یک داده‌ی sample داریم با پارامتر $\hat{\theta}$ و می‌خواهیم با تخمینگرهای نقطه‌ای یک تخمین ارزیابی $\hat{\theta}$ برای θ ارائه بدیم.

$X_1 \dots X_{10}$

$$\hat{\mu}_1 = X_1 \qquad E(\hat{\mu}_1) = \mu$$

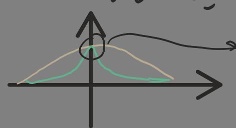
$$\hat{\mu}_2 = \bar{X} = \frac{X_1 + X_2 + \dots + X_{10}}{10} \qquad E(\hat{\mu}_2) = \mu$$

کدوم یکی از $\hat{\mu}_1$ و $\hat{\mu}_2$ بهتره؟

$\hat{\mu}_2$ چون واریانسش کمتره

معیارهای ارزیابی تخمینگرهای نقطه‌ای

Var: معیارهایی برای تخمینگرها داریم که دقت آن‌ها را با هم مقایسه می‌کنه. یکی از این معیارها Var هست. Var کمتر باشه بهتر.



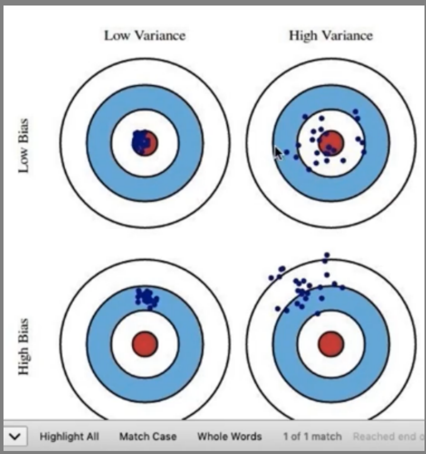
هر دو داری یک مقدار برمی‌گردونن اما سبزه بهتره چون Var کمتری داره

Bias: از رابطه‌ی روبه‌رو استفاده می‌کنیم و اگر که

$$B(\hat{\theta}) = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta$$

$B(\hat{\theta}) = 0$ یعنی $E(\hat{\theta}) = \theta$ اون موقع تخمینگر رو نااریب یا (Unbiased Estimator) می‌نامیم

← ما به دنبال پیدا کردن این قبیل تخمینگرها هستیم.



باتوجه به این تصاویر ما به دنبال تخمینگری هستیم که بایاس کم و واریانس کمی داشته باشه.

$$E((\hat{\theta}-\theta)^2)$$

Mean Squared Error : متوسط مربع خطا. به ماشین لرنینگ به شدت کاربرد داره و از رابطه‌ی روبه‌رو به دست می‌آید.

$$Var(X) = E(X^2) - E(X)^2 \longrightarrow Var(\hat{\theta}-\theta) = E((\hat{\theta}-\theta)^2) - E(\hat{\theta}-\theta)^2$$

$$= Var(\hat{\theta}) = MSE(\hat{\theta}) - Bias^2(\hat{\theta})$$

به رابطه‌ی طلایی 3 تا معیار ارزیابی می‌رسیم.

$$MSE = Var(\hat{\theta}) + Bias(\hat{\theta})^2$$

مثال: sample $= X_1, X_2, \dots, X_{10}$, $\hat{\mu}_2 = \bar{X}$, $\hat{\mu}_1 = X_1$

از دیدگاه Bias با هم تفاوتی ندارند.

از دیدگاه Var :

$$Var(\hat{\mu}_1) = \sigma^2$$

$$Var(\hat{\mu}_2) = \frac{\sigma^2}{n}$$ → بهتره چون واریانس کمتری داره

$$MSE(\hat{\mu}_1) = \sigma^2 - 0 = \sigma^2$$

$$MSE(\hat{\mu}_2) = \frac{\sigma^2}{n} - 0 = \frac{\sigma^2}{n}$$

پس با توجه به برابر بودن Bias ها و واریانس کمتر، MSE کمتر برای $\hat{\mu}_2$ ⇐ $\hat{\mu}_2$ تخمینگر بهتری هست.

آیا می‌توان Estimator داشت که همزمان $Var=0$ باشد اما Bias داشته باشد؟ آره مثلاً $\hat{\mu}_3 = 19$

چون عدد ثابته واریانس نداره و $Bias = 19 - \mu$ ن

معیار سازگاری (Consistency)

اگر $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n$ یک دنباله از تخمین‌گرها باشند

$$\forall \varepsilon > 0 \quad \lim_{n \to \infty} \mathbb{P}\left(|\hat{\theta}_n - \theta| \geq \varepsilon\right) = 0$$

یعنی به حالتی که تعداد تخمین‌گرهام زیاد باشه و فاصلش از معیار اصلی θ به سمت ۰ بره اون موقع اون دنباله از تخمین‌گرها رو سازگار می‌نامیم.

مثلاً

$$\hat{\theta}_1 = X, \quad \hat{\theta}_2 = \frac{X_1 + X_2}{2}, \quad \dots, \quad \hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

اینا فقط تئوری هستن و در دنیای واقعی ما نمی‌تونیم بی نهایت تا سمپل بگیریم بلکه روی کاغذ ثابت می‌کنیم روش سمپل گیری ما و تخمین‌گرها مون حالتی که به صفر میل کنه اون موقع سازگار است.

مثال تخمین‌گر $\hat{\theta}_n = \frac{1}{n}\sum_i X_i + \frac{1}{n}$ را بررسی کنید.

$$E(\hat{\theta}_n - \mu) = \frac{1}{n} = B(\hat{\theta}_n) \Rightarrow$$ unbiased نیست

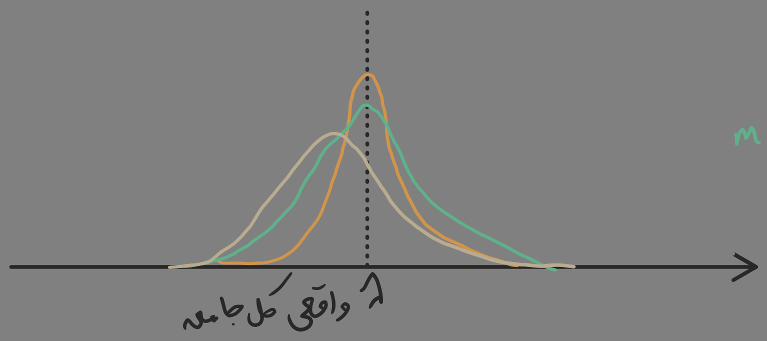
if $n \to \infty$ then $\frac{1}{n} \to 0$ ~ طبق قانون اعداد بزرگ ثابت میشه که میل میکنه به سمت صفر $\Rightarrow$ سازگار است.

تخمین‌گری داریم که unbiased باشد در عین حال سازگار هم نباشه.

$$\hat{\theta}_1 = X_1, \quad \hat{\theta}_2 = X_1, \quad \dots, \quad \hat{\theta}_n = X_1 \quad \text{or} \quad \hat{\theta}_n = X_n$$

بایاس نداره و سازگار هم نیست.

مقایسه‌ی چند تخمینگر



Var — هم بایاس داره و هم واریانس زیادی داره

mean(5 by 5) — واریانس یکم زیاده اما بایاس نیست.

mean — واریانس خیلی کم و بایاس هم نیست

چرا Var , biased هست و استیل Unbiased اش چی هست؟

$$\bar{S}^2 = \frac{1}{n}\sum (X_i-\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n}\bar{X}^2 \cancel{-\frac{2}{n}\sum_{i=1}^{n} X_i\bar{X}}$$

چون $\sum X_i = n\bar{X}$

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

$$(Var(X) = E(X^2) - E(X)^2)$$

$$E(\bar{S}^2) = \frac{1}{n}\sum E(X_i^2) - \frac{1}{n}\sum E(\bar{X}^2) = \frac{1}{n}\sum E(X_i^2) - \frac{\sigma^2}{n} - \mu^2$$

$$E(X_i^2) = \underbrace{Var(X_i)}_{\sigma^2} + \underbrace{E(X_i)^2}_{\mu^2}$$

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

چون برابر σ^2 نشد پس biased است

برای Unbiased کردن ← کل ما بارو در $\frac{n}{n-1}$ ضرب کنیم که درست کار کنه

$$\hat{\sigma}^2 = \frac{n}{n-1}\bar{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

الان این دیگه نااریب هست (Unbiased)

این همه معیارهای تخمینگر خوب پیدا کردیم. حالا اصلاً چطوری یک تخمینگر بسازیم؟ روش ساختن تخمینگر!!

تخمینگر بیشینه درست نمایی (Maximum Likelihood Estimator)

کاربرد شکلین در یادگیری ماشین

یک sample داریم که از توزیع خاصی (مثلاً Poisson) و می‌خواهیم یک θ بدست بیاریم که احتمال دیدن داده‌های sample با این θ بیشینه بشه.

$$\mathbb{P}(X_1, X_2, \dots, X_n ; \theta) \rightsquigarrow \text{Max}$$

فرض کنید مثلاً θ رو ۱۰ ، ۱۰۰ بدست آوردید. با این وضعیت طبیعتاً اگه sample روی پواسون باشه قطعاً احتمال اینکه θ برابر ۱۰ باشه واقعاً بیشتره پس $\theta = 10$ رو گزارش می‌دیم.

نکته: هم برای توزیع های گسسته و هم برای توزیع های پیوسته کار میکنه

تابع Likelihood برای گسسته پیوسته

$$\begin{cases} L(\theta \mid X) = PMF_{\theta}(X) & \text{Discrete} \\ L(\theta \mid X) = PDF_{\theta}(X) & \text{Contin} \end{cases}$$