Subject:
Date

احتمال : مدل های ریاضی دارای مجهولات و فرضیات

- روش حل مسئله
  - سهشودی
  - تحلیلی ← مدل سازی
    - مدارس
    - فیزیکی
    - منطقی
    - شبیه سازی
    - ریاضی
    - مفهومی
    - آوایی و سیستمی

مدل : مابه ازای واقعیت برای درک بهتر واقعیت

- مدل سازی ریاضی
  - عدم قطعیت ← تصادفی ←
    - فضای احتمال
    - متغیر تصادفی
    - بردار تصادفی
    - معادلات دیفرانسیل تصادفی
    - فرآیند تصادفی
    - گراف های تصادفی

هرگاه بوسیله ی ابزارهای ریاضی بتوانیم واقعیت را بیان کنیم به آن مدل می گوییم.

فضای احتمال ← آزمایش تصادفی

متغیر تصادفی ← کمیت تصادفی در آزمایش تصادفی

بردار تصادفی ← چند کمیت تصادفی در آزمایش تصادفی

آزمایش : مجموعه ای از قدم های مشخص و قابل تکرار

P4PCO

آزمایش قطعی : آزمایشی که نتیجه‌ی آن قابل پیش‌بینی باشد

آزمایش غیرقطعی : آزمایشی که نتیجه‌ی آن غیرقابل پیش‌بینی باشد.

آزمایش تصادفی : آزمایش غیرقطعی که فضای رخداد آن ثابت باشد.

ساده‌ترین آزمایش تصادفی : برنولی

برنولی → شکست 0

برنولی → پیروزی 1

آزمایش دوجمله‌ای : اگر آزمایش های برنولی مستقل و یکسان را n بار تکرار کنیم.

مثال : پرتاب یک سکه 10 بار

کمیت های ثابت : n ، احتمال شیر آمدن در یک آزمایش ، احتمال خط آمدن در یک آزمایش

کمیت های متغیر : تعداد شیر آمدن در n بار آزمایش ، تعداد دنباله ها ،
طول بلندترین یا کوتاه‌ترین دنباله .

متغیر تصادفی : $X : \Omega \xrightarrow[\text{اندازه پذیر بورل}]{} \mathbb{R}$

بردار های تصادفی $\equiv$ توابع توزیع توأم

متغیر تصادفی : قد یک فرد ، وزن یک فرد

بردار تصادفی : تالیس از قد و وزن یک فرد

معادلات دیفرانسیل تصادفی : اگر ضرایب یک معادله‌ی دیفرانسیل ثابت نباشد .

گراف های تصادفی : در شبکه‌ی مجازی هر فرد را یک گره و بررسی کنیم که آیا این افراد باهم در ارتباط هستند یا نه .

آمار : در مدل های احتمال یک سری مجهولات وجود دارد که می خواهیم این مجهولات را پیدا کنیم . یعنی ابتدا باید با استفاده از مدل های تصادفی پیدا کنیم که هر اتفاق چه توزیعی دارد بعد مجهولات آن توزیع را پیدا کنیم .

احتمال : نوع توزیع را همراه با مجهولات آن به ما می دهند .

توابع توزیع گسسته : برنولی ـ دوجمله ای ـ هندسی ـ دوجمله ای منفی ـ پواسون ـ فوق هندسی ـ یکنواخت گسسته

توابع توزیع پیوسته : یکنواخت ـ نرمال ـ نمایی ـ گاما (ارلنگ ـ کای دو) ـ کوشی .

مدل سازی تصادفی
- ۱ ـ آمار
- ۲ ـ احتمال
- ۳ ـ شبیه سازی

Subject:
Date

اهمیت آمار و احتمال
- ← پیش نیاز
- ← کاربرد : وجود انسان و عدم قطعیت
- ← دیدگاه
- ← کنکور ارشد
- ← پژوهش

- آمار
  - مقدمات
    - مروری کوتاه بر احتمال
    - توزیع های جدید T و F
    - نمونه ی تصادفی
  - برآورد
    - نقطه ای
    - بازه ای
  - آزمون
  - پیش بینی

برآورد کردن : تخمین زدن پارامترهای تابع توزیع.

آزمون کردن : تست کردن فرضیات.

پیش بینی کردن : استفاده ی روابط بیان متغیرهای تصادفی برای پیدا کردن پارامترهای تصادفی مجهول.

$(\Omega, \mathcal{F}, P)$ فضای نمونه

$X : \Omega \xrightarrow[\text{اندازه پذیر بورل}]{} \mathbb{R}$ متغیر تصادفی

1) CDF : تابع توزیع تجمعی $F_X(x) = P_r(X \leq x)$

خواص CDF : غیر نزولی
حد در بی نهایت برابر با 1
از راست پیوسته.

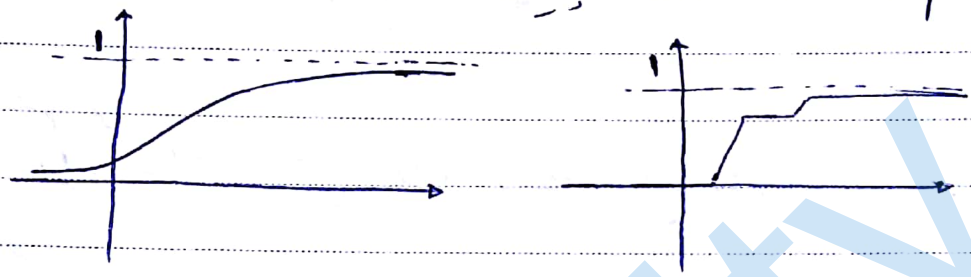
CDF پیوسته : از راست و چپ پیوسته.

CDF گسسته : پلکانی و [illegible].

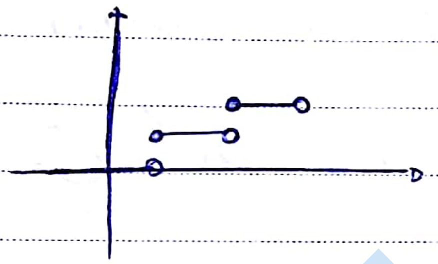
Subject:
Date

اگر تابع توزیع تجمعی نقاط ناپیوستگی (پلکانی) باشد نمی‌توانیم نتیجه بگیریم که حتماً گسسته است بلکه باید تعداد پله‌ها شمارا باشد.

CDF : هیچ کدام از خواص گسسته و پیوسته را ندارد.

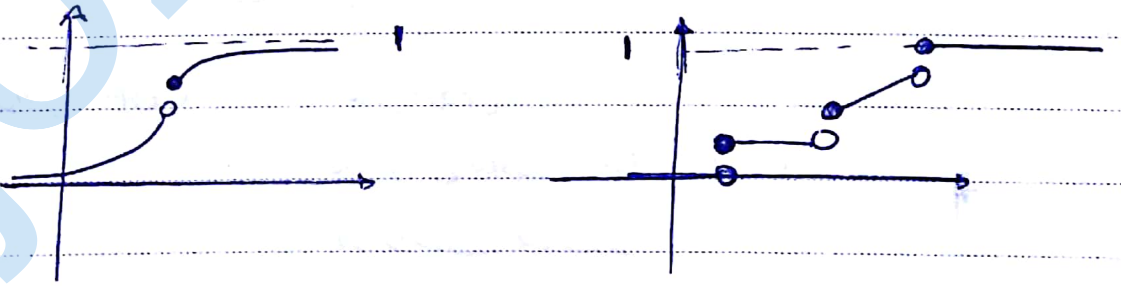


CDF پیوسته



CDF گسسته

تابع توزیع کانتور یک تابع توزیع پیوسته است ولی CDF آن پله‌ای است اما تعداد پله‌های آن شمارا نیست. و چون گسسته‌ها همیشه pmf دارند ولی پیوسته‌ها همیشه pdf ندارند چون کانتور تابع توزیع ندارد پس پیوسته است و pdf آن موجود نیست.



CDF مخلوط

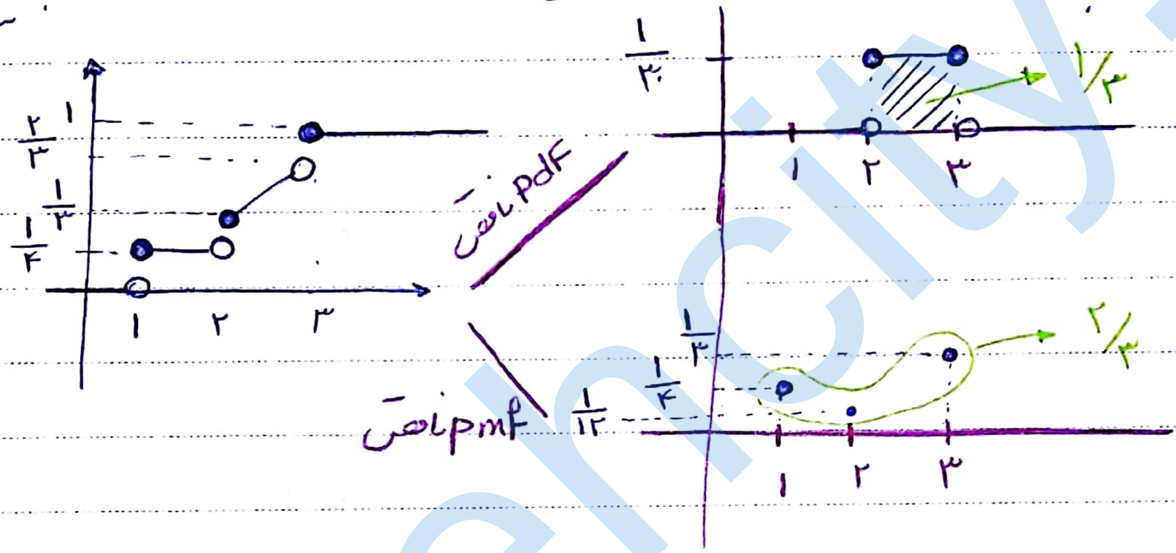
Date

خصوصیات pmf و PDF : ۱) نامنفی

۲) جمع همه‌ی مقادیر آن‌ها برابر ۱.

pmf ناقص و PDF ناقص : نامنفی هستند اما جمع همه‌ی مقادیر آن ۱ نیست.

CDF مخلوط برای زمانی است که pmf ناقص و PDF ناقص داشته باشیم.



$1/3 + 2/3 = 1$

$$E(x) = \underbrace{\int_2^3 \frac{1}{3} x \, dx}_{\text{امید ریاضی پیوسته}} + \underbrace{1 \times \frac{1}{4} + 2 \times \frac{1}{12} + 3 \times \frac{1}{3}}_{\text{امید ریاضی گسسته}}$$

اگر CDF مخلوط داشته باشیم باید PDF ناقص را پیدا کنیم (که مشتق CDF است..) و pmf ناقص را هم پیدا کنیم (که فاصله‌ی بین پرش‌هاست..) اگر مجموع این دو ۱ باشد می‌توانیم مسأله را حل کنیم.

Scanned by CamScanner

Subject
Date

جلسه سوم :

$\lambda > 0$

$X \sim Exp(\lambda)$

$f_X(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$

$F_X(x) = (1 - e^{-\lambda x}) I_{(0,\infty)}(x)$

$E(x) = \frac{1}{\lambda}$

$Var(x) = \frac{1}{\lambda^2}$

$M_X(t) = \frac{\lambda}{\lambda - t} \qquad t < \lambda$

خاصیت عدم حافظه

$X - t \mid X > t \sim X$

$$P_r(X - t < s \mid X > t) = \frac{P_r(t < X < t+s)}{P_r(X > t)}$$

$$= \frac{1 - e^{-\lambda(t+s)} - (1 - e^{-t\lambda})}{1 - (1 - e^{-\lambda t})} = 1 - e^{-\lambda s} = P_r(X < s)$$

در متون آماری به جای $\frac{1}{\lambda}$ ، β قرار می دهیم :

$\frac{1}{\lambda} \sim \beta$

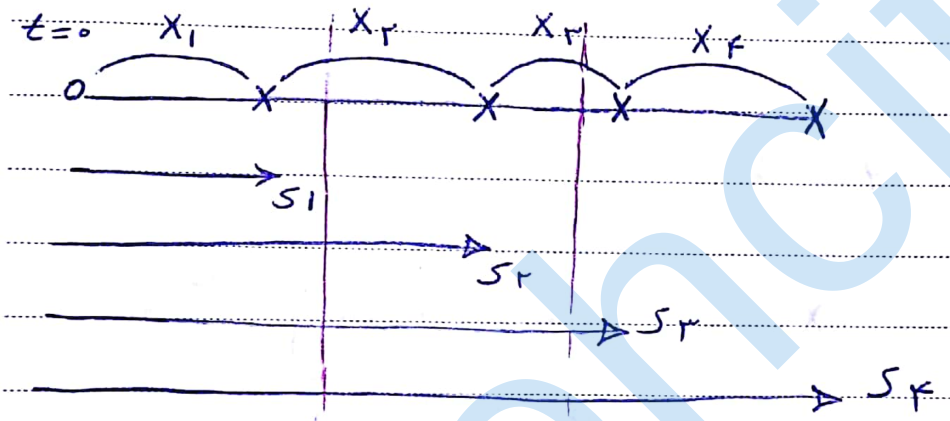
$\Rightarrow E(x) = \beta \qquad Var(x) = \beta^2$

فرآیند تصادفی: به مجموعه‌ای از متغیرهای تصادفی، فرآیند تصادفی می‌گویند.

$$\{X_t\}_{t \in T}$$

○ بردارهای تصادفی، تعداد متناهی از متغیرهای تصادفی است اما فرآیندهای تصادفی تعداد نامتناهی شمارا یا ناشمارا از متغیرهای تصادفی است.

فرآیند تجدید:



زمان رخداد X_1

زمان رخداد X_2

$S_1 = X_1$

$S_2 = X_1 + X_2$

$S_3 = X_1 + X_2 + X_3$

$S_4 = X_1 + X_2 + X_3 + X_4$

$N(t_1)$ $N(t_2)$

در این جا ۳ متغیر تصادفی داریم:

$\{X_i\}_{i \in \mathbb{N}}$: زمان بین دو واقعه

$\{S_n\}_{n \in \mathbb{N}}$: زمان وقوع هر رخداد

$\{N(t)\}_{t \in \mathbb{R}^{\geq 0}}$: تعداد پدیده‌هایی که تا زمان t اُم رخ می‌دهند.

(نامتناهی ناشمارا)

Subject
Date

اگر X_i ها iid (مستقل و هم توزیع) باشند در این صورت به این فرآیند، فرآیند تجدید می گویند.

$$\circ \; P_r(S_n > t) = P_r(N(t) \leq n-1)$$

خاصیت فرآیند تجدید

(t)

اثبات: چون S_n زمان رخ داد n اُمین پدیده است پس در این صورت حتماً n اُمین پدیده بعد از زمان t رخ داده است پس در این صورت تا زمان t حداکثر $n-1$ پدیده رخ داده است.

فرآیند پواسون:

اگر X_i ها iid باشند و توزیع نمائی داشته باشند در این صورت فرآیند تجدید تبدیل به فرآیند پواسون با توزیع λ می شود.

$$X_i \sim Exp(\lambda)$$

$$N(t) \sim pos(\lambda t)$$

$$S_n \sim G(n, \lambda)$$

• اگر زمان بین وقوع هریک از رخدادها از توزیع نمایی پیروی کند در این صورت تعداد رخدادهای اتفاق افتاده تا زمان t از توزیع پواسون با پارامتر λt پیروی می‌کند.

$Y \sim G(n, \lambda)$ (توزیع گاما)

$$F_Y(y) = P_r(Y \leqslant y) = P_r(S_n \leqslant y) = 1 - P_r(S_n > y)$$

$$= 1 - P_r(N(y) \leqslant n-1) = 1 - P_r(Pos(\lambda y) \leqslant n-1) =$$

$$= \begin{cases} 1 - \sum_{i=0}^{n-1} \frac{(\lambda y)^i e^{-\lambda y}}{i!} & y \geqslant 0 \\ 0 & y < 0 \end{cases}$$

Subject
Date

توزیع گاما : $X \sim G(\alpha, \lambda)$ $\qquad \frac{1}{\lambda} \sim \beta$

$$f_X(x) = \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x} I_{(0,\infty)}(x)$$

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$\Gamma(n) = (n-1)!$ $\qquad \Gamma(1/2) = \sqrt{\pi}$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

نمایی $\alpha = 1$ $\qquad \lambda = \lambda$ $\qquad Exp(\lambda)$

ارلنگ $\alpha = n$ $\qquad \lambda = \lambda$ $\qquad Er(n, \lambda)$

کای دو $\alpha = \frac{n}{2}$ $\qquad \lambda = \frac{1}{2}$ $\qquad \chi^2_n$

$E(X) = \alpha/\lambda$

$Var(X) = \alpha/\lambda^2$

تابع مولد گشتاور : $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$ $\qquad t < \lambda$

$X \sim G(\alpha, \lambda) \qquad \alpha + r > 0$

$$E(x^r) = \frac{\Gamma(\alpha + r)}{\lambda^r \Gamma(\alpha)}$$

r لزوماً صحیح نیست.

خواص توزیع گاما:

$$X \sim G(\alpha, \lambda), \quad Y \sim G(\beta, \lambda)$$

مستقل

1) $X + Y \sim G(\alpha + \beta, \lambda)$

مستقل

2) $\dfrac{X}{X+Y} \sim Beta(\alpha, \beta)$

مستقل

3) $\dfrac{X}{Y} \sim c F_{2\alpha + 2\beta}$

اگر فرض کنید $Z_1 = \dfrac{X}{X+Y}$, $Z_2 = X/Y$ در این صورت

Z_1 تابعی از Z_2 خواهد بود مثلاً اگر داشته باشیم:

$$P_r(Z_1 \leq a, \; Z_2 \leq b)$$

$$= P_r(f(Z_2) \leq a, \; Z_2 \leq b)$$

$$Z_1 = \frac{Z_2}{Z_2 + 1} \leq a \longrightarrow (1 - a) Z_2 \leq a$$

$$Z_2 \leq \frac{a}{1 - a}$$

$$\rightarrow P_r\left(Z_2 \leqslant \frac{a}{1-a}\ ,\ Z_2 \leqslant b\right)$$

$$= P_r\left(Z_2 \leqslant \min\left\{\frac{a}{1-a}\ ,\ b\right\}\right)$$

و می‌توانیم Z_2 را هم برحسب Z_1 نوشت بطوری که

$$Z_2 = \frac{Z_1}{1-Z_1}$$

مثال : اگر $X \sim \chi^2_2$ ، $Y \sim \chi^2_2$ مستقل از هم باشند

$$P_r\left(X+Y > 4 \mid 2X > X+Y\right)$$

$$= P_r\left(X+Y > 4 \mid \frac{X}{X+Y} > \frac{1}{2}\right)$$

چون می‌دانیم X و Y توزیع گاما دارند و مستقل هستند پس $X+Y$

و $\frac{X}{X+Y}$ هم از هم مستقل هستند.

$$\rightarrow P_r(X+Y > 4) = P_r\left(\chi^2_4 > 4\right) = \sum_{i=0}^{2-1} \frac{(4 \times 1/2)^i e^{-4 \times 1/2}}{i!}$$

۴) اگر $X \sim G(\alpha, \lambda)$, $c > 0$

$$\rightarrow cX \sim G(\alpha, \lambda/c)$$

۵) اگر X_i ها iid , $X_i \sim Exp(\lambda)$

$$\rightarrow \bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \qquad X_i \sim G(1, \lambda)$$

$$\sum_{i=1}^{n} X_i \sim G(n, \lambda) \quad , \quad \frac{1}{n} > 0$$

$$\rightarrow \frac{1}{n} \sum_{i=1}^{n} X_i \sim G\left(n, \frac{\lambda}{\frac{1}{n}}\right)$$

$$\rightarrow \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim G(n, n\lambda)$$

۶) $2\lambda\, G(n, \lambda) = G(n, \frac{1}{2}) \sim \chi^2_{2n}$

۷) $X_i \sim Exp(\lambda)$, X_i : iid

$$\bar{X} \sim G(n, n\lambda)$$

$$\frac{1}{2n\lambda} \times 2n\lambda\, G(n, n\lambda) = \frac{1}{2n\lambda} G(n, \frac{1}{2})$$

$$= \frac{1}{2n\lambda} \chi^2_{2n}$$

$$\rightarrow \bar{X} \sim \frac{1}{2n\lambda} \chi^2_{2n}$$

توزیع نرمال :

$$X \sim N(\mu, \sigma^2)$$

$$\mu \in \mathbb{R} \,, \quad \sigma > 0$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$Z \sim N(0,1) \rightarrow f_Z(x) = \phi(x)$$

$$ \hookrightarrow F_Z(x) = \Phi(x)$$

$$E(X) = \mu$$

$$E(X^2) = \mu^2 \sigma^2$$

$$Var(X) = \sigma^2 \qquad E\left((X-\mu)^r\right) = \begin{cases} 0 & r = 2k+1 \\ ? & r = 2k \end{cases}$$

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \qquad t \in \mathbb{R}$$

F

Subject
Date

- $E(x^3) = ?$

$E((x-\mu)^3) = 0$

$E(x^3) - 3\mu E(x^2) + 3\mu^2 E(x) - \mu^3 = 0$

$\longrightarrow E(x^3) = 3\mu\sigma^2 + \mu^3$

خواص توزیع نرمال :

$$\text{مستقل} \begin{cases} X \sim N(\mu_1, \sigma_1^2) \\ Y \sim N(\mu_2, \sigma_2^2) \end{cases}$$

1) $\alpha X + \beta Y + \gamma \sim N(\alpha\mu_1 + \beta\mu_1 + \gamma, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2)$

$$\frac{X-\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \sim N\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{1}{\sigma^2} \times \sigma^2\right)$$

$$= N(0,1) \implies \frac{X-\mu}{\sigma} \sim Z \sim N(0,1)$$

2) $X_i \sim N(\mu_i, \sigma_i^2)$ مستقل

$$\sum_{i=1}^{n} \alpha_i X_i \sim N\left(\sum_{i=1}^{n} \alpha_i \mu_i, \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2\right)$$

Subject
Date

۳) $X_i \sim N(\mu, \sigma^2)$ iid

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

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

از خاصیت دوم اگر $\alpha_i = \frac{1}{n}$ و $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$ باشد.

۴) $M_{\alpha X + \beta Y + \gamma}(t) = M_X(\alpha t) M_Y(\beta t) e^{t\gamma}$

اگر X و Y مستقل.

۵) $M_{\sum_{i=1}^{n} \alpha_i X_i}(t) = \prod_{i=1}^{n} M_{X_i}(\alpha t_i)$

۶) می‌دانیم: $\frac{X_i - \mu_i}{\sigma_i} \sim Z$

و چون $Z^2 \sim \chi_1^2$ پس $\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_1^2$

۷) $(X_i - \mu_i)^2 = ?$

$\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_1^2 = G(1/2, 1/2)$

$(X_i - \mu_i)^2 = \sigma^2 G(1/2, 1/2) = G(1/2, \frac{1}{2\sigma^2})$

Subject

Date

۸) $X_i \sim N(\mu_i, \sigma_i^2)$

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2_n$$

$X_i \sim N(\mu, \sigma^2)$ مستقل

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2_n$$

$$\sum_{i=1}^{n} (X_i - \mu)^2 \sim \sigma^2 \chi^2_n = G\left(n/2, \frac{1}{2\sigma^2} \right)$$

Date

$X_i \sim N(\mu_i, \sigma_i^2)$ مستقل

مهم : $\sum_{i=1}^{n} \alpha_i X_i \sim N\left(\sum_{i=1}^{n} \alpha_i \mu_i, \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2\right)$

$\hookrightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\circledast \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_n^2$

$\hookrightarrow \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi_n^2$

واریانس نمونه : $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

$\hookrightarrow S^2 \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{\frac{2\sigma^2}{n-1}}\right)$

اگر X_i ها iid باشند , $\text{Cov}(X_i - \bar{X}, \bar{X}) = 0$

درنتیجه $E(X_i^2) < \infty$

$X_i , \quad i = 1, \dots, n \quad$ iid

$E(X_i) < \infty \Rightarrow X_i - \bar{X}, \bar{X}$: ناهمبسته

Subject
Date

$\bar{X}$ و بردار $(X_1-\bar{X}, X_2-\bar{X}, \dots, X_n-\bar{X})$ مستقل است $\Longleftrightarrow$

X_i ها از توزیع نرمال پیروی می کند.

اثبات: اگر $\bar{X}$ از بردار بالا مستقل باشد از تک تک $X_i-\bar{X}$

هم مستقل است وقتی از همه مستقل باشد، از همه ناهمبسته هم

است پس $Cov(X_i-\bar{X}, \bar{X})=0$ که این خاصیت، خصیصه توزیع

نرمال است.

دو متغیر که از هم ناهمبسته باشند لزوماً مستقل نیستند.

مثال: $Cov(Z, Z^2)=0$ درحالی که Z و Z^2 ناهمبسته هستند اما از

هم مستقل نیستند.

$$f_{\bar{X},(X_1-\bar{X},\dots,X_n-\bar{X})}(a,b_1,\dots,b_n)=f_{\bar{X}}(a)f_{X_i-\bar{X}}(b_1,\dots,b_n)$$

• اگر $\bar{X}$ از $X_i-\bar{X}$ مستقل باشد از هر تابع اندازه پذیری از آن ها

مستقل خواهد بود. مثلاً:

$$Cov(\bar{X}, (X_1-\bar{X})(X_2-\bar{X}))=0$$

• می دانیم $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ تابعی از $x_i - \bar{x}$ است

پس $\bar{x}$ از s^2 مستقل است و $Cov(\bar{x}, s^2) = 0$

(به شرطی که $x_i \sim N(\mu, \sigma^2)$ و iid باشند.)

* اگر $X + Y = Z \Longrightarrow Y \sim \chi^2_{n-m}$

X, Y مستقل

$X \to \chi^2_m \qquad Z \to \chi^2_n \qquad m < n$

اثبات:

$$M_{x+y}(t) = M_z(t)$$

$$M_x(t)\, M_y(t) = M_z(t)$$

$$(1-2t)^{\frac{-m}{2}} M_y(t) = (1-2t)^{\frac{-n}{2}}$$

$$\longrightarrow M_y(t) = (1-2t)^{\frac{-(n-m)}{2}} \Rightarrow Y \sim \chi^2_{n-m}$$

• $X + Y = Z$

X, Y مستقل، Z نرمال

توزیع کوشی

مستقل
$$\left. \begin{array}{l} X \sim N(0, \sigma^2) \\ Y \sim N(0, \sigma^2) \end{array} \right\} \longrightarrow \frac{X}{Y} \sim \varphi(0,1)$$

$$\frac{X}{|Y|} \sim \varphi(0,1)$$

$$\varphi(0,1) = \frac{1}{\pi(1+x^2)}$$

یک اتحاد ریاضی وجود دارد بصورتی $\bar{x}$ میانگین x_i هاست :

$$\sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - a)^2$$

اثبات :

$$\sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} \left((x_i - \bar{x})^2 + (\bar{x} - a)^2 + 2(x_i - \bar{x})(\bar{x} - a)\right)$$

$$+2\sum_{i=1}^{n} (x_i - \bar{x})(\bar{x} - a)$$

$$= \sum_{i=1}^{n} (x_i - \cancel{\bar{x}} + \cancel{\bar{x}} - a)^2$$

$$\longrightarrow 2\sum_{i=1}^{n} (x_i - \bar{x})(\bar{x} - a) = 0 \longrightarrow \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

اگر در اتحاد بالا به جای a ، μ بگذاریم و دو طرف رابطه را در $\frac{1}{\sigma^2}$ ضرب کنیم در این صورت :

$$\sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 + \left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

χ^2_n مستقل χ^2_1

$$\longrightarrow \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma}\right)^2 \sim \chi^2_{n-1}$$

مهم :

Subject
Date

$$S'^2 = \frac{\sum (x_i-\mu)^2}{n}$$

$$, \quad \underbrace{\frac{\sum_{i=1}^{n} (x_i-\mu)^2}{\sigma^2}}_{\frac{nS'^2}{\sigma^2}} \sim \chi^2_n$$

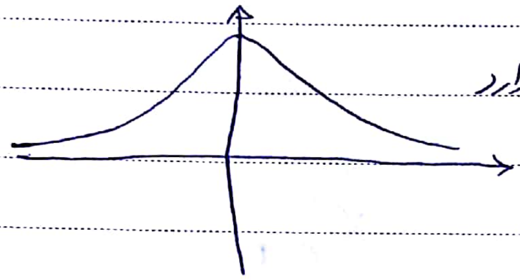
$$\longrightarrow \frac{\sum (x_i-\bar{x})^2}{\sigma^2} = \frac{S^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

Subject
Date

توزیع T :

$T_n \quad n\in\mathbb{N}$ (n می‌تواند هر عدد مثبتی باشد)

$$f_{T_n}(t)=\frac{\Gamma(n+1)}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad t\in\mathbb{R}$$



شبیه توزیع نرمال استاندارد

CDF فقط در حالت $n=1$ به راحتی بدست می‌آید.

$$F_{T_n}(t)\Big|_{n=1}=\int_{-\infty}^{+\infty}\frac{\Gamma(2)}{\Gamma(1/2)\sqrt{\pi}}(1+t^2)^{-1}$$

$$=\frac{-1}{\pi}\,\text{arc Cotg}(t)$$

برای $n=1$ توزیع T همان توزیع کوشی نرمال است

1) $T_1 \sim \varphi(0,1)$

2) $T_n \xrightarrow{D(L)} Z \qquad \lim_{n\to\infty} F_{T_n}(t)=\Phi(t)$

3) $T_n \sim \dfrac{X}{\sqrt{\frac{Y}{n}}} \qquad \left.\begin{matrix} X\sim N(0,1) \\ Y\sim \chi^2_n \end{matrix}\right\}$ مستقل

$T_n \sim \dfrac{Z}{\sqrt{\chi^2_n/n}} \qquad Z$: همیشه نرمال استاندارد

Z و χ^2_n : باید از هم مستقل باشند

Subject
Date

توزیع T ، حول صفر تابع مولد گشتاور ندارد بنابراین تمامی توابع گشتاور ها خام

حتی k اُمش هم برای هر k این وجود ندارد .

$$E(T_n) = E\left(\frac{Z}{\sqrt{\frac{\chi^2_n}{n}}}\right) = E(Z)\,E\left(\frac{\sqrt{n}}{\sqrt{\chi^2_n}}\right)$$

بر اساس فرمول $\underbrace{E(G(x)\,H(y))}_{\text{مستقل}} \overset{\text{هر دو متناهی باشند}}{=\!=} E(G(x))\,E(H(y))$

$E(Z)=0$ اگر $E\left(\frac{\sqrt{n}}{\sqrt{\chi^2_n}}\right)$ موجود باشد رابطه ی بالا درست است .

$$E\left(\frac{\sqrt{n}}{\sqrt{\chi^2_n}}\right) = \sqrt{n}\,E\left((\chi^2_n)^{-1/2}\right)$$

$$E(X^r) = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)\,\lambda^r} \qquad X \sim G(\alpha, r)$$

$$r+\alpha>0 \longrightarrow r>-\alpha$$

$$\chi^2_n \sim G(n/2, 1/2) \qquad r=-1/2$$

$$n/2 - 1/2 > 0 \longrightarrow n>1$$

$$E\left(\frac{\sqrt{n}}{\sqrt{\chi^2_n}}\right) = \sqrt{n}\,\frac{\Gamma(n/2 - 1/2)}{\Gamma(n/2)\,(1/2)^{-1/2}} \qquad n>1$$

$$E(T_n) = 0$$

Subject
Date

$n>1 \quad E(T_n)=0$

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

$= E\left(\frac{Z^2 n}{\chi^2_n}\right) = \frac{n}{n-2} \quad n>2$

$E(Z^2)=1$

$Var(T_n) = \frac{n}{n-2} \quad n>2$

اگر iid ~ X_i ها و Z_1 و Z_2 نرمال استاندارد مستقل باشند.

$T_1 \sim \varphi(0,1) = \frac{Z_1}{|Z_2|} \quad n=1$

یادآوری : $X_i \sim iid$

* $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim Z$

* $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

* $\frac{nS'^2}{\sigma^2} \sim \chi^2_n$

* $\bar{X}$ و S^2 همیشه از هم مستقل هستند.

* $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim T_{n-1}$

Subject
Date

اثبات :

$$\frac{\underbrace{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}_{Z}}{\sqrt{\frac{\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}}}{n-1}}} \sim \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim T_{n-1}$$

ساده شده‌ی کسر بالا همان $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ است که چون $\bar{X}$ و S^2 از هم

مستقل هستند Z ، χ^2_{n-1} هم از هم مستقل هستند

? $\quad \frac{c\,(X_{n+1}-\mu)}{\sqrt{\sum_{i=1}^{n}(X_i-\mu)^2}} \sim T_n \qquad X_1,\dots,X_{n+1} \sim NID(\mu,\sigma^2)$

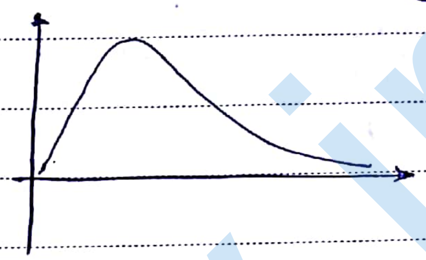
$\frac{c\,(X_1-X_2)}{\sqrt{(X_1+X_2)^2}} \sim T_1 \qquad X_1, X_2 \sim N(0,\sigma^2)$

توزیع F :

$$F_{m,n} = \frac{\frac{\chi^2_m}{m}}{\frac{\chi^2_n}{n}}$$ مستقل

درجه

نمایش توزیع F :



$$* \quad (T_n)^2 \sim \frac{Z^2}{\frac{\chi^2_n}{n}} \sim \frac{\chi^2_1}{\frac{\chi^2_n}{n}} \sim F_{1,n}$$

فرض کنید $X \sim G(\alpha, \lambda_1)$ ، $Y \sim G(\beta, \lambda_2)$ ثابت کنید در این صورت $c\,\frac{X}{Y} \sim F_{2\alpha, 2\beta}$.

$$\frac{X}{Y} = \frac{\lambda_2}{\lambda_1} \times \frac{\lambda_1 X}{\lambda_2 Y} = \frac{\lambda_2}{\lambda_1} \frac{\lambda_1}{\lambda_2} \frac{G(\alpha, \lambda_1)}{G(\beta, \lambda_2)} =$$

$$= \frac{\lambda_2}{\lambda_1} \frac{G(\alpha, 1)}{G(\beta, 1)} = \frac{2\lambda_2}{2\lambda_1} \frac{G(\frac{2\alpha}{2}, 1/2)}{G(2\beta/2, 1/2)} =$$

$$\frac{2\lambda_2}{2\lambda_1} \frac{\chi^2_{2\alpha}}{\chi^2_{2\beta}} \times \alpha/\beta \times \beta/\alpha = \frac{\lambda_2}{\lambda_1} \alpha/\beta \; F_{2\alpha, 2\beta}$$

Subject

Date

$$f_{F_{m,n}}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}} I_{(0,\infty)}(x)$$

دو حالت خاص مهم :

$$* \quad f_{F_{2,2}}(x) = \frac{1}{(1+x)^2} \qquad x>0$$

$$* \quad f_{F_{4,4}}(x) = \frac{6x}{(1+x)^4} \qquad x>0$$

ویژگی مهم توزیع F : مستقل

$$\left.\begin{array}{l} X \sim \chi^2_m \\ Y \sim \chi^2_n \end{array}\right\} \Rightarrow \frac{X/m}{Y/n} \sim F_{m,n}$$

$$\Rightarrow \frac{1}{F_{m,n}} \sim F_{n,m}$$

$$* \quad T_n^2 \sim F_{1,n}$$

$$* \quad m F_{m,n} \xrightarrow[n\to\infty]{L/D} \chi^2_m$$

زیرا اگر $X_i \sim \chi^2_1$ ، $\mu = 1$ ، $Y = \frac{\sum_{i=1}^n X_i}{n}$

$$\sum_{i=1}^n X_i / n \xrightarrow[P]{a.s.} \mu = 1$$

PAPCO

Subject
Date

مثال $P_r(F_{1,1} \overset{a>0}{>} a) = P_r(T_1^2 > a) = P_r(\underbrace{T_1>\sqrt{a}}_{\varphi(0,1)}) + P_r(T_1<-\sqrt{a})$

مثال $P_r(\underbrace{F_{4,200}}_{\frac{200-2}{200}} < 2) \approx P_r(\underbrace{\chi^2_4}_{G(2,\frac{1}{2})} < 8) = 1 - \sum_{i=0} \frac{(8\times \frac{1}{2})^i e^{-8\times \frac{1}{2}}}{i!}$

$$E_{F_{m,n}}(X) = \frac{n}{n-2}$$

مثال:

مستقل $\begin{cases} X \sim \chi^2_4 \\ Y \sim G(2,8) \end{cases}$ $\quad P_r(X>Y) = ?$

راه حل عادی : (مسیر طولانی)

$$P_r(X>Y) = \iint_{x>y} f_X(x) f_Y(y)\, dx\, dy$$

$$\frac{X}{Y} = \frac{\chi^2_4}{\frac{16}{16}G(2,8)} = \frac{\chi^2_4}{\frac{1}{16}G(2,\frac{1}{2})} = \frac{\chi^2_4}{\frac{1}{16}\chi^2_4} = 16\left(\frac{\chi^2_4/4}{\chi^2_4/4}\right) = 16F_{4,4}$$

$$P_r\left(\frac{X}{Y} > 1\right) = P_r(16F_{4,4} > 1) = P_r(F_{4,4} > 1/16)$$

$$= \int_{1/16}^{+\infty} \frac{6x}{(1+x)^4}\, dx$$

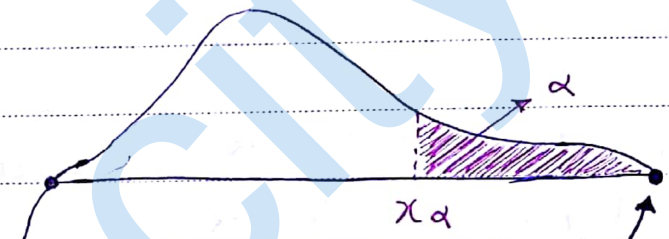
Subject
Date

$$= \int_{1/16}^{+\infty} \frac{6x}{(1+x)^4}\,dx = 6\left(\frac{(1+x)^{-3+1}}{-3+1} - \frac{(1+x)^{-4+1}}{-4+1} \right)\Bigg|_{1/16}^{+\infty}$$

$$6\left(\frac{x}{(1+x)^4}\right) = 6\left(\frac{x+1}{(1+x)^4} - \frac{1}{(1+x)^4}\right) = 6\left(\frac{1}{(1+x)^3} - \frac{1}{(1+x)^4}\right)$$

$$= P_r(X > Y) = 6\left(\frac{1}{2\left(1+\frac{1}{16}\right)^2} - \frac{1}{3\left(1+\frac{1}{16}\right)^3}\right)$$

X صدک بالای α اُم : x_α



$Z \longrightarrow z_\alpha$

$T_n \longrightarrow t_{n,\alpha}$

$F_{m,n} \longrightarrow f_{m,n,\alpha}$

$\chi^2_n \longrightarrow \chi^2_{n,\alpha}$

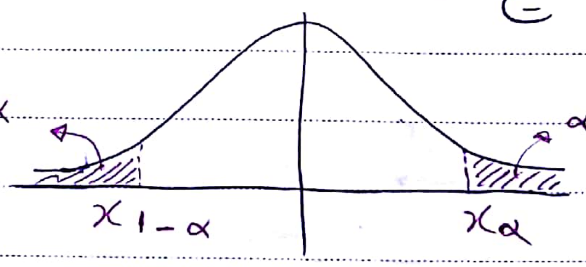
$x_0 = \text{ess sup}(X) = REP(X)$

$x_1 = \text{ess inf}(X) = LEP(X)$

$$F_X(x_\alpha) = P_r(X \leq x_\alpha) = 1 - P_r(X > x_\alpha) = 1 - \alpha$$

$F_X(x_\alpha) = 1 - \alpha$ برای توزیع های متقارن

$F^{-1}(1-\alpha) = x_\alpha$ α



$$f_{n,m,\alpha} = \frac{1}{f_{m,n,1-\alpha}}$$

Subject

Date

$$Z_\alpha = -Z_{1-\alpha}$$

$$t_{n,\alpha} = -t_{n,1-\alpha}$$

$$f_{m,n,\alpha} = \frac{1}{f_{n,m,1-\alpha}}$$

$$\alpha = P_r(F_{m,n} \geq f_{m,n,\alpha}) = P_r\left(F_{m,n} \geq \frac{1}{f_{n,m,1-\alpha}}\right)$$

$$= P_r\left(f_{n,m,1-\alpha} \geq \frac{1}{F_{m,n}}\right) = P_r(f_{n,m,1-\alpha} \geq F_{n,m})$$

$$= \alpha \qquad \Longrightarrow f_{m,n,\alpha} = \frac{1}{f_{n,m,1-\alpha}}$$

اگر $\begin{cases} P_r(x \geq a) = \alpha \\ P_r(x \geq b) = \alpha \end{cases} \Longrightarrow a = b$

مستقل

$X_i \sim N(\mu_1, \sigma_1^2)$ $\qquad\qquad$ $Y_i \sim N(\mu_2, \sigma_2^2)$

X_1 $\qquad$ Y_1

X_2 $\qquad$ Y_2

$\vdots$ $\qquad$ $\vdots$

X_n $\qquad$ Y_n

$\bar{X}$ $\qquad$ $\bar{Y}$

S_x^2 $\qquad$ S_y^2

$$\Longrightarrow \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} \sim Z \qquad \left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim Z\right)$$

Subject
Date

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\frac{\sigma_2^2}{\sigma_1^2}\,\frac{S_x^2}{S_y^2} = \frac{\dfrac{S_x^2(n_1-1)}{\sigma_1^2(n_1-1)}}{\dfrac{S_y^2(n_2-1)}{\sigma_2^2(n_2-1)}} = \frac{\dfrac{1}{(n_1-1)}\,\dfrac{S_x^2(n_1-1)}{\sigma_1^2}}{\dfrac{1}{(n_2-1)}\,\dfrac{S_y^2(n_2-1)}{\sigma_2^2}}$$

$$= \frac{\dfrac{\chi^2_{n_1-1}}{n_1-1}}{\dfrac{\chi^2_{n_2-1}}{n_2-1}} \sim F_{n_1-1,\, n_2-1}$$

$$\frac{\sigma_2^2}{\sigma_1^2}\,\frac{S_x^2}{S_y^2} \sim F_{n_1-1,\, n_2-1}$$

$$\left(\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim T_{n-1}\right) \qquad \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim T_{n_1+n_2-2}$$

$$S_p^2 = \frac{S_x^2(n_1-1)+S_y^2(n_2-1)}{n_1+n_2-2}$$

$$\frac{\underbrace{\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}_{\to Z}}{\sqrt{\dfrac{\underbrace{\dfrac{S_p^2(n_1+n_2-2)}{\sigma^2}}_{\to \chi^2_{n_1+n_2-2}}}{n_1+n_2-2}}} \sim T_{n_1+n_2-2}$$

مستقل

PAPCO

Subject
Date

$$\frac{S_p^2\,(n_1+n_2-2)}{\sigma^2} = \frac{S_x^2\,(n_1-1)}{\sigma^2} + \frac{S_y^2\,(n_2-1)}{\sigma^2}$$

$$= \chi^2_{n_1-1} + \chi^2_{n_2-1} = \chi^2_{n_1+n_2-2}$$

نمونه‌گیری به صورت زوجی :

$$P_1 = X_1 - Y_1 \longleftarrow (X_1, Y_1)$$
$$\vdots \qquad\qquad \vdots$$
$$P_n = X_n - Y_n \longleftarrow (X_n, Y_n)$$

اگر $(X, Y) = N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$ ⟵ توزیع نرمال بیضوی

توزیع نرمال بیضوی :

$$X = \begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix} \sim N\left(\mu = \begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}, \Sigma = \begin{pmatrix}\sigma_1^2 & \cdots & \sigma_{1n}\\ \sigma_{12} & \ddots & \vdots\\ \vdots & & \sigma_n^2\\ \sigma_{1n} & \cdots & \end{pmatrix}\right)$$

$$\longrightarrow X \sim N(\mu, \Sigma)$$

$\Sigma \rightsquigarrow$ ماتریس کوواریانس $\qquad \Sigma = (\sigma_{ij})_{i,j=1}^{n}$

$\mu \in \mathbb{R}^n \qquad \sigma_{ij} = Cov(X_i, X_j)$

Σ : یک ماتریس متقارن نیمه معین مثبت است .

P4PCO

تعریف ماتریس متقارن نیمه معین مثبت :

psd

$\forall x \in \mathbb{R}^n \qquad x^T \Sigma x \geq 0$

///

$\lambda_1(\Sigma), \dots, \lambda_n(\Sigma) \geq 0$

یعنی تمام مقدارهای ویژه ماتریس متقارن نامنفی است .

pd

← اگر تمام مقدارهای ویژه ماتریس متقارن مثبت باشد، ماتریس متقارن معین مثبت است .

قضیه : هر ماتریس کوواریانس (Σ) نیمه معین مثبت و متقارن است .

$$Var\left(\sum_{i=1}^{n} x_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \, Cov(X_i, X_j)$$

$$= x^T \Sigma x \geq 0$$

pd $\quad \Sigma > 0 \qquad f_X(x) = \frac{1}{(\pi)^{n/2} \, |\Sigma|^{n/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$

Σ: pd توزیع نرمال چند متغیره pdf

اگر $n=1 \qquad f_X(x) = \frac{1}{\sqrt{\pi}\,\sigma_1} e^{\frac{-(x-\mu_1)^2}{\sigma_1^2}}$

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N(\mu, \Sigma) \qquad \mu \in \mathbb{R}^n$$

$$e^A = 1 + \frac{A}{1!} + \frac{A^2}{2!} + \dots$$

Subject
Date

اگر $(X, Y) \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$

$$\alpha X + \beta Y + \gamma \sim N(\alpha\mu_1 + \beta\mu_2 + \gamma, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2 + 2\alpha\beta\sigma_{12})$$

و هم چنین برای توزیع نرمال توام بیضوی

اگر المان های غیر قطری ماتریس Cov همگی صفر باشد $\Leftarrow$ X_i ها همه از هم مستقل

(برای بقیه توزیع ها از استقلال X_i ها می توان نتیجه گرفت که Cov صفر است نه برعکس.)

اگر $(X, Y) = \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$

$$D_i = X_i - Y_i$$

$$\Longrightarrow \bar{D} = \bar{X} - \bar{Y}$$

$$\rightarrow D = X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$$

$$\rightarrow \frac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}}} \sim Z$$

$$\rightarrow S_D^2 = \frac{\sum_{i=0}^{n} (D_i - \bar{D})^2}{n-1}$$

$$\rightarrow S_{xy} = \frac{\sum_{i=0}^{n} (x_i - \bar{X})(y_i - \bar{Y})}{n-1}$$

$$\rightarrow \frac{\bar{D} - (\mu_1 - \mu_2)}{\frac{S_D}{\sqrt{n}}} \sim T_{n-1}$$

Subject:
Date

جزوه‌ی TA : x_1 ، x_2 ، ... ، x_n ← کوچک / مستقل / هم توزیع

کوچکترین بین x_1 تا x_n — $x_{(1)}$ آماره ترتیبی اول

دومین کوچکترین بین $x_1 ... x_n$ — $x_{(2)}$

⋮

j اُمین کوچکترین بین $x_1 ... x_n$ — $x_{(j)}$

⋮

بزرگترین = n اُمین کوچکترین بین $x_1 ... x_n$ — $x_{(n)}$ آماره ترتیبی n اُم

هر تابع از x_1 ، ... ، x_n ← آماره

آماره های ترتیبی:
- CDF
- pdf
- حالت خاص جابجایی ← x_1
- حالت خاص جابجه تکراری نیست

فرض کنید X_1 ، X_2 ، ... ، X_n متغیرهای تصادفی پیوسته، مستقل و هم توزیع با تابع چگالی مشترک f و تابع توزیع تجمعی مشترک F هستند. در آمار و مهندسی به هر تابعی از متغیرهای تصادفی مستقل و هم توزیع $x_1, x_2, ..., x_n$ یک آماره می‌گویند.

و یک نوع از آماره ها بسیار مهم و کاربردی در علم آمار، آماره های ترتیبی می باشند.

$X_1 = 8:20$ → $X_{(1)} = 8:10$

$X_2 = 8:10$ $X_{(2)} = 8:15$

$X_3 = 8:15$ $X_{(3)} = 8:20$

تابع توزیع تجمعی آماره‌های ترتیبی :

$$F_{X_{(k)}}(a) = Pr(X_{(k)} \le a) =$$

Pr(K تا از n آماره کمتر مساوی a باشند) + ... + Pr(n تا از n تا کمتر مساوی a باشند)

$$\binom{n}{k}(F(a))^k(1-F(a))^{n-k} + \dots + \binom{n}{n}(F(a))^n \underbrace{(1-F(a))^0}$$

$$= \sum_{j=k}^{n}\binom{n}{j}(F(a))^j(1-F(a))^{n-j}$$

n تا از آماره‌ها کمتر از a

۰ تا از آماره‌ها بیشتر از a

تابع چگالی آماره‌های ترتیبی :

روش اول : $\dfrac{dF_{X_{(k)}}(a)}{da} = f_{X_{(k)}}(a)$

روش دوم : $\dfrac{dF(x)}{dx} = f(x) \longrightarrow dF(x) = f(x)\,dx$

$$f_{X_{(k)}}(a)\cancel{da} = \binom{n}{k-1}\times\binom{n-k+1}{1}\binom{n-k}{n-k}\,\underbrace{f(a)\cancel{da}}\,(F(a))^{k-1}(1-F(a))^{n-k}$$

k-1 تا از آماره‌ها کمتر از a

دقیقا یکی از آماره‌ها برابر a باشد (آماره‌ی k ام)

n-k تا از آماره‌ها بیشتر از a

احتمال اینکه آماره k ام برابر a

$$f_{X_{(k)}}(a) = \binom{n}{k-1}\binom{n-k+1}{1}\binom{n-k}{n-k} f(a)(F(a))^{k-1}(1-F(a))^{n-k}$$

pdf

$$F_{X_{(k)}}(a) = \sum_{j=k}^{n}\binom{n}{j}(F(a))^j(1-F(a))^{n-j} \longrightarrow \text{CDF}$$

Subject:
Date

اگر $X_1, \dots, X_n$ متغیرهای تصادفی مستقل و یکنواخت در بازه‌ی $(0, 1)$ باشند آن گاه توزیع آماره‌ی مرتبه‌ی i اُم یعنی $X_{(i)}$ از توزیع $Beta(i, n-i+1)$ پیروی می‌کند.

مستقل

$X_1, \dots, X_n \sim Uniform(0,1)$

$X_{(i)} \sim Beta(i, n-i+1)$

$$f_{X(K)}(a) = \binom{n}{i-1}\binom{n-i+1}{1}\binom{n-i}{n-i} f(a)\left(F(a)\right)^{i-1}\left(1-F(a)\right)^{n-i}$$

pdf یکنواخت در بازه‌ی $(0, 1)$ CDF یکنواخت در بازه $(0,1)$

$$f_{X(i)}(a) = \underbrace{\frac{n!}{(i-1)!(n-i)!}}_{\text{ثابت}} \times 1 \times (a)^{i-1}(1-a)^{n-i} \sim Beta(i, n-i+1)$$

$0 \leq a \leq 1$

۳ لامپ را درنظر بگیرید که طول عمر هرکدام نمایی با پارامتر λ است. ۳ لامپ را بصورت همزمان روشن می‌کنیم.

الف) امید ریاضی مدت زمان تا سوختن اولین لامپ

ب) 〃 〃 〃 دومین لامپ

ج) 〃 〃 〃 سومین لامپ

Subject:
Date

الف)

$$f_{X(k)}(a) = \frac{n!}{(k-1)!(n-k)!}\,\underbrace{\lambda e^{-\lambda a}}_{\text{pdf توزیع نمائی } \lambda\ (f(a))}\,\underbrace{(1-e^{-\lambda a})}_{\text{CDF توزیع نمائی } \lambda}^{i-1}\,(e^{-\lambda a})^{n-i}$$

$$k=1 \qquad f_{X_{(1)}}(a) = \frac{n!}{(n-1)!}\,\lambda e^{-\lambda a}\,e^{-\lambda a(n-1)}$$

$$= n\lambda e^{-\lambda a} \sim \exp(n\lambda)$$

امید ریاضی الف $\quad E(X) = \frac{1}{n\lambda} = \frac{1}{۳\lambda}$

ب) $\quad E(X_1 + X_۲) = \frac{1}{۳\lambda} + \frac{1}{۲\lambda}$

X_1: اولین لامپ از ۳ تا

$X_۲$: اولین لامپ از ۲ تای باقی مانده

مدلسازی تصادفی

فضای احتمال — متغیر تصادفی

آزمایش تصادفی — کمیت تصادفی ← تابع توزیع F_θ

واقعیت ⟷ اعتبارسنجی ⟷ [مدلسازی ⟷ صحت‌سنجی]

X	F_θ
X_1	x_1
X_2	x_2
$\vdots$	$\vdots$
X_n	x_n
نمونه نظری	نمونه عملی

وقتی که هنوز نمونه‌گیری را شروع نکردیم هر نمونه توزیعش X_1 تا X_n است و هنوز احتمالاتی نداریم به خاطر اینکه هم توزیع با X است.

اما هنگامی که نمونه‌گیری که x_1 تا x_n نمونه‌ها واقعی هستند.

اگر $X_1, \dots, X_n \sim iid$ باشند و هم توزیع باشند در این صورت یک نمونه تصادفی از آن توزیع هستند.

آمار:
- توصیفی: تحلیل اعداد به معنای کار با میانگین و میانه و واریانس و ... سر و کار داشتن با نمونه‌ی عملی (مقدار)
- استنباطی: قبل از نمونه‌گیری عملی، نمونه‌گیری نظری انجام می‌دهیم و با نمونه‌های واقعی سر و کار داریم.

Subject:
Date

به هر چیزی که قابلیت نمونه گیری از آن وجود داشته باشد، جامعه می گویند.

محدودیت های تجربی درباره‌ی واقعیت + دامنه‌ی ریاضیات $\rightarrow$ $\theta \in \Theta$

مثلاً اگر قد ایرانیان توزیع $N(\mu, \sigma^2)$ باشد مطمئن هستم که $\mu < 2m$

یعنی محدودیت تجربی ما این است که میانگین از ۲ متر بیشتر نیست

و دامنه‌ی ریاضیات این است که $\mu \in \mathbb{R}$ باشد.

* آماره: یک تابع اندازه پذیر لبگ از X_1 تا X_n که تنها پارامتر مجهول که <u>می تواند</u> در آن باشد n هست.

اگر n تعداد نمونه باشد

X_1 : آماره $\quad$ $X_2 + X_1$ = آماره $\quad$ $e^{X_1} + X_2$ = آماره

$nX_1 + Y_1$: آماره $\quad$ $\sum X_i$, $\bar{X}$ = آماره

اما اگر θ مجهول باشد θX_1 آماره نیست.

چون در آماره مجهول جز n نداریم و در ساختار آن هیچ پارامتر مجهول دیگری نیست

بنابراین وقتی نمونه مشاهده شود، آماره هم مشاهده می شود.

Subject:
Date

مثال) $x_1 = 4 \quad x_2 = -4 \quad x_3 = 1 \quad x_4 = 5$

$\rightarrow n = 4$

$X_1 \rightarrow x_1$

$\vdots$

$X_n \rightarrow x_4$

$n = 4 \rightarrow \bar{x} = \frac{4-4+1+5}{4}$

$\rightarrow \bar{x} = 3/2$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$ جمع کردن آماره ها

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$ (μ : مجهول   آماره نیست : S'^2)

$Y_1 = \min\{X_1 \dots X_n\} = X_{(1)}$ آماره های ترتیبی

$Y_n = \max\{X_1 \dots X_n\} = X_{(n)}$

$D = Y_n - Y_1 = X_{(n)} - X_{(1)}$ دامنه

$MD = \frac{Y_1 + Y_n}{2} = \frac{X_{(1)} + X_{(n)}}{2}$ میان دامنه

$$M = \begin{cases} Y_{k+1} & n = 2k+1 \\ \frac{Y_k + Y_{k+1}}{2} & n = 2k \end{cases}$$ میانه

$$\overline{X^k} = \frac{\sum_{i=1}^{n} X_i^k}{n}$$ گشتاور خام مرتبه k ام

P4PCO

Subject:
Date

گشتاور مرکزی مرتبه k اُم $\quad CM_k^* = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^k}{n-k}$

برآوردگر $\begin{cases} \text{نقطه‌ای} \\ \text{بازه‌ای} \end{cases}$

برآوردگر های نقطه‌ای : T را یک برآوردگر نقطه‌ای برای پارامتر θ گویند

$T(X_1 \dots X_n)$

اگر T یک آماره باشد و توزیع آن به θ وابسته باشد، مشاهده‌ی T را

یک برآورد برای θ گوییم.

برای مثال اگر $\theta = 3$ باشد می‌توانیم $T = \bar{X}$ بگیریم.

اگر $\bar{x} = 1.5$ باشد $\bar{x}$ را یک مشاهده برای $\bar{X}$ و یک

برآورد برای ۳ می‌گوییم.

مقایسه برآوردگرهای نقطه‌ای :

(برآوردگر بهتر ⟵ خطای کمتر)

$err_T(\theta) = |T - \theta|$ خطای برآوردگر

$E(|T-\theta|)$ میانگین خطای برآوردگر

$E(|T-\theta|^2) = MSE_T(\theta)$ میانگین مربع خطاها

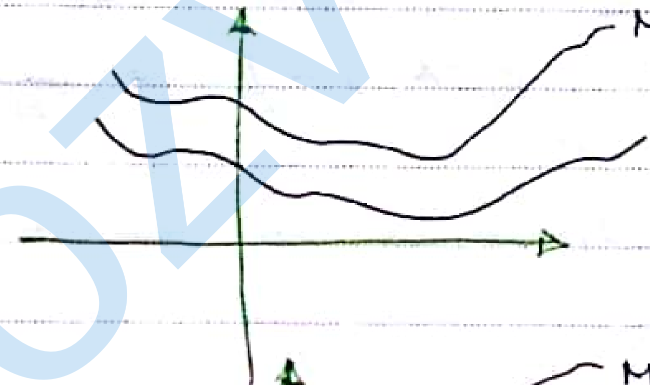
(means of square errors)

T_1 از T_2 بهتر است اگر

$$MSE_{T_1}(\theta) \leq MSE_{T_2}(\theta) \quad \forall \theta \in \Theta$$

$$T_2 \text{ ضریب کارایی } T_1 \text{ نسبت به } = \frac{MSE_{T_2}(\theta)}{MSE_{T_1}(\theta)} \begin{cases} >1 & T_1 \text{ بهتر از } T_2 \\ =1 & \text{کارایی یکسان} \\ <1 & T_2 \text{ بهتر از } T_1 \end{cases}$$

* این روابط باید به ازای تمام مقادیر θ درست باشد.



همیشه $MSE_{T_1} < MSE_{T_2}$

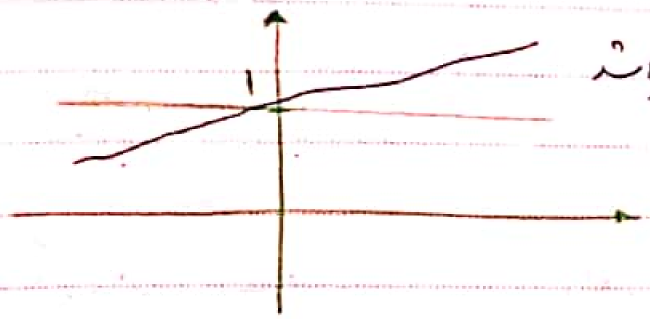
پس T_1 کاراتر است.

* به ازای تمام θ درست نیست

یعنی نتوانیم نتیجه بگیریم

اگر به ازای همه‌ی θ ها T_1 کاراتر از T_2 باشد

* نمودار ضریب کارایی T_1 نسبت به T_2



اگر به ازای همه‌ی θ ها MSE_T ها یکنواخت نباشد

* نمودار ضریب کارایی T_1 نسبت به T_2

مثال) اگر $\theta = \mu$ باشد و $T_1 = X_1$ و $T_2 = \bar{X}$ کدام یک کاراتر است؟

$$MSE_{T_1}(\theta) = E\left((X_1-\mu)^2\right) = E\left(X_1^2+\mu^2-2X_1\mu\right)$$

$$= E(X_1^2) + \mu^2 - 2E(X_1)\mu = E(X^2) - \mu^2 = \sigma^2$$

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

$$MSE_{T_2}(\theta) = E\left((\bar{X}-\mu)^2\right) = E\left(\bar{X}^2+\mu^2-2\bar{X}\mu\right)$$

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

$MSE_{T_1} > MSE_{T_2} \longrightarrow$ T_2 کاراتر از T_1

$$MSE_T(\theta) = E\left((T-\theta)^۲\right) = E\left(\left(T - E(T) + E(T) - \theta\right)^۲\right)$$

$$= E\left(\left(T-E(T)\right)^۲\right) + E\left(\left(E(T)-\theta\right)^۲\right) + ۲E\left(۲\left(T-E(T)\right)\left(E(T)-\theta\right)\right)$$

$$= Var(T) + \underset{\downarrow}{\left(E(T)-\theta\right)^۲} + ۲\underbrace{E\left(T-E(T)\right)}_{۰}\left(E(T)-\theta\right)$$

$$\left(B_T(\theta)\right)^۲$$

$$MSE_T(\theta) = Var(T) + \left(B_T(\theta)\right)^۲$$

$\left(B_T(\theta)\right)^۲$ به میزان اریبی θ گفته می‌شود و بصورت زیر تعریف می‌شود

$\alpha'\sigma^۲ - \sigma^۲$

$$\left(B_T(\theta)\right)^۲ = \left(E(T)-\theta\right)^۲$$

* برآوردگر نااریب : به برآوردگری گفته می‌شود که ضریب اریب بودن آن صفر باشد یعنی $B_T(\theta) = ۰$ یا $MSE_T(\theta) = Var(T)$

$$\begin{cases} E(T) = \theta & \forall \theta \in \Theta \\ \quad \Updownarrow & \\ B_T(\theta) = ۰ & \forall \theta \in \Theta \end{cases}$$

و $MSE_T(\theta) = Var(T)$

برآوردگر مجانباً نااریب :

$$\begin{cases} \mathcal{L}\ E(T) = \theta & \forall \theta \in \Theta \\ \mathcal{L}\ B_T(\theta) = ۰ & \forall \theta \in \Theta \end{cases}$$

Subject:
Date:

* هر برآوردگر ناریب، مجانباً هم ناریب است ولی هر برآوردگر مجانباً ناریب، برآوردگر ناریب نیست.

مثال) فرض کنید $X_i \sim NID(\mu, \sigma^2)$ و $T_\alpha = \alpha S^2$ ، $\alpha > 0$ یک کلاسی از برآوردگرها باشند، برای بهترین برآوردگر α را محاسبه کنید.

$\theta = \sigma^2$ است.

ناریب $\left\{ \begin{array}{l} T_1 = S^2 \\ E(S^2) = \sigma^2 \end{array} \right\}$

$$MSE_{T_\alpha}(\theta) = E\left((T_\alpha - \sigma^2)^2\right)$$

$$Var(T_\alpha) + (B_T(\alpha))^2 = Var(\alpha S^2) + (E(\alpha S^2) - \sigma^2)^2$$

$$\alpha^2 Var(S^2) + (\alpha - 1)^2 \sigma^4$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = Var(\chi^2_{n-1}) = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma^4} Var(S^2) = 2(n-1) \rightarrow Var(S^2) = \frac{2\sigma^4}{(n-1)}$$

$$MSE_{T_\alpha}(\theta) = \frac{2\alpha^2\sigma^4}{n-1} + (\alpha-1)^2\sigma^4 = \sigma^4\left(\frac{2\alpha^2}{n-1} + (\alpha-1)^2\right)$$

$$h(\alpha) = \frac{2\alpha^2}{n-1} + (\alpha-1)^2$$

این تابع محدب است زیرا دوبار مشتق آن نامنفی است ✓

PAPCO

Subject:
Date

نقطه‌ی Mini تابع محدب جایی است که مشتق اول آن صفر بشود.

$$\frac{4\alpha}{n-1}+2(\alpha-1)=0$$

$$\frac{2\alpha}{n-1}=1-\alpha$$

$$2\alpha+(n-1)\alpha=n-1 \longrightarrow (n+1)\alpha=n-1$$

$$\alpha^*=\frac{n-1}{n+1}$$

اگر $T_{\alpha^*}=\frac{n-1}{n+1}S^2$ اگر $T_1=S^2$

مجانباً ناریب ناریب

$$MSE_{T_1}(\sigma^2) \geqslant MSE_{T_\alpha}(\sigma^2)$$

* لزوماً برآوردگری که ناریب است کاراتر از مجانباً ناریب نیست

Subject:
Date

← آیا اریبی یک برآوردگر رفع شدنی است؟

برای مثال یک برآوردگر بصورت $E(T) = a\theta + b$ در این صورت این برآوردگر

اریب است اما با یک تبدیل آمنی می‌تواند به یک برآوردگر نااریب تبدیل بشود.

$T' = \frac{T-b}{a}$ ، $E(T') = \theta$ برآوردگر نااریب است.

$a=1 \quad b=0$

$a = \frac{n-1}{n} \quad b=0$

یا به طور مثال اگر $T = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$ و $E(T) = \frac{n-1}{n}\sigma^2$ در این

صورت $T = S^2$ که $T' = \frac{1}{\frac{n-1}{n}}$ $E(T' = S^2) = \sigma^2$ که نااریب

$a=1 \quad b=0$

است.

ممکن است یک برآوردگر مجانباً اریب را هیچ وقت نتوانیم به برآوردگر نااریب تبدیل کنیم.

⇐ برآورد سازگار: وقتی که $n \to \infty$ می‌رود برآوردگر نااریب می‌شود.

برآوردگر مجانباً نااریب: وقتی که $n \to \infty$ می‌رود برآوردگر اریب می‌شود.

برآوردگر سازگار $\begin{cases} \text{سازگار ضعیف} \\ \text{سازگار قوی} \end{cases}$

سازگار ضعیف : $\lim\limits_{n\to\infty} P_r\{|T-\theta|<\varepsilon\}=1$ $\quad \forall\theta\in\Theta$

سازگار قوی : $\lim\limits_{n\to\infty} MSE_T(\theta)=0$ $\quad \varepsilon>0 \quad \forall\theta\in\Theta$

یعنی وقتی که $n\to\infty$ می رود خطا صفر میشود که انتظارش را داریم زیرا

$n=\infty$ یعنی سرشماری پس اگر سرشماری کنیم نباید خطایی داشته باشیم.

* اگر یک برآوردگر سازگار قوی باشد ۱ و ۲ را نتیجه می‌گیریم

۱) سازگار ضعیف هم هست.

۲) مجانبا نااریب هم هست + $\lim\limits_{n\to\infty} var(T)=0$

اثبات ۲ : $MSE_T(\theta)=var(T)+(B_T(\theta))^2$ پس

$$\lim_{n\to\infty} MSE_T(\theta)=\lim_{n\to\infty} var(T)+\lim_{n\to\infty}\{(B_T(\theta))^2\}=0$$

$$\to \lim_{n\to\infty} var(T)=0$$

$$\to \lim_{n\to\infty} (B_T(\theta))^2=0$$ پس مجانبا نااریب

اثبات ۱ :

$$Pr\{|T-\theta|<\varepsilon\} = 1 - Pr\{|T-\theta|>\varepsilon\}$$

$$Pr\{|T-\theta|>\varepsilon\} = Pr\{\underbrace{|T-\theta|^2}_{Y\geq 0} > \underbrace{\varepsilon^2}_{a}\} \overset{\text{مارکوف}}{\leq} \frac{E(Y)}{a}$$

$$E(Y) = E((T-\theta)^2) = MSE_T(\theta)$$

$$0 \leq Pr\{|T-\theta|>\varepsilon\} \leq \frac{MSE_T(\theta)}{a} \to 0$$

لا یه (سازگاره)

$$\longrightarrow \lim_{n\to\infty} Pr\{|T-\theta|>\varepsilon\} = 0$$

$$\longrightarrow \lim_{n\to\infty} Pr\{|T-\theta|<\varepsilon\} = 1$$ سازگار ضعیف

$$Pr(Y \geq a) \leq \frac{E(Y)}{a}$$ نامساوی مارکوف

۱) $Pr(Y\geq 0) = 1$

۲) $a>0$

۳) $E(Y)<\infty$

Subject:
Date

آیا برآوردگر بهینه برای θ در حالت کلی می‌توان پیدا نمود؟ .

(بر اساس معیار MSE)

$$\forall T \quad \text{with} \quad var(T) < \infty$$

$$MSE_{T^*}(\theta) \leqslant MSE(\theta) \qquad \forall \theta \in \Theta$$

برآوردگر بهینه

o جامعه نابدیهی : یعنی متغیر تصادفی حداقل دو مقدار را با احتمال غیرصفر بگیرد.

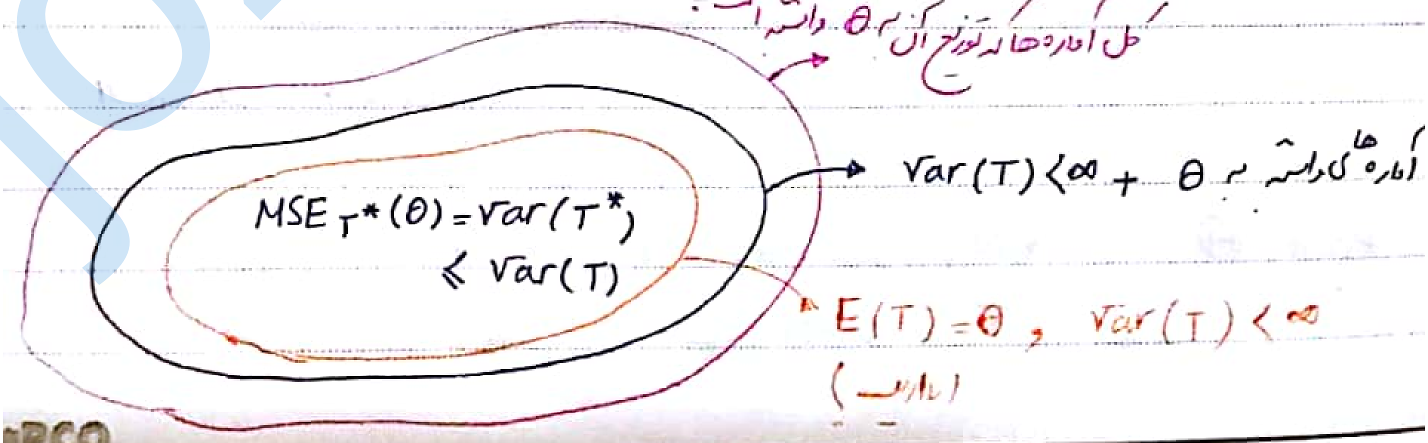
$$X \sim U(\theta - 1/2 \,,\, \theta + 1/2) \qquad \theta \in \mathbb{R}$$

o جامعه‌ی بدیهی : جامعه ای که همواره یک مقدار را می‌گیرد.

$$X \sim N(1 \,,\, 0)$$

یعنی چون میانگین 1 است و واریانس صفر است همه‌ی داده ها برابر با 1 است.

می‌توانیم مثالی بیاوریم که متغیر بهینه ، برآوردگر نااریب باشد؟



Scanned by CamScanner

اگر T^* برآوردگر بهینه باشد می‌توانیم آن در حلقه‌ی دوم که $Var(T)<\infty$ است

پیدا می‌کنیم که البته ما حلقه را کوچکتر می‌کنیم و دیگر برآوردگرهای ناریب به دنبال

آن می‌گردیم (هرچند ممکن است برآوردگر بهینه، برآوردگر ناریب نباشد.) اما با

این کار برآوردگر ناریب بهینه را پیدا می‌کنیم.

$$E(T^*)=\theta \qquad \forall T': \ E(T')=\theta$$

$$MSE_{T^*}(\theta) \leq MSE_{T'}(\theta)$$

$$\parallel$$

$$Var(T^*) \leq Var(T') \qquad \forall\theta\in\Theta$$

به برآوردگر بهینه‌ی ناریب UMVUE می‌گویند

uniform — min var — ناریب — برآوردگر (estimator)

بدست آوردن UMVUE :

تحت شرایط مشخص (که از قضیه معلوم است) اگر T برآوردگر ناریبی باشد

داریم : $Var(T) \geq CR(n,\theta)$ (کران پایین کرامر رائو)

استفاده از کران پایین :

اگر $\forall\theta\in\Theta \quad Var(T)=CR(n,\theta)$

آنگاه $\longrightarrow T = UMVUE$

Subject:
Date

یعنی به ازای هر برآوردگر نااریب T یک کران پایین وجود دارد به نام

$CR(n,\theta)$ (کرامر رائو) (که یعنی $Var(T)$ همیشه از CR بزرگتر

است.) در این صورت اگر T' را طوری بیابیم که نااریب باشد و

$Var(T')=CR(n,\theta)$ باشد در این صورت T' برآوردگر یکنواخت بهینه

$(UMVUE)$ است.

$$CR(n;\theta)=\frac{1}{nE\left(\left(\frac{\partial \ln(f_x(x;\theta))}{\partial\theta}\right)^2\right)}$$

یا

$$CR(n,\theta)=\frac{1}{nE\left(-\frac{\partial^2 \ln(f_x(x;\theta))}{\partial^2\theta}\right)}$$

f_x، pdf یا pmf توزیع است و شرایطی که قبلاً گفتیم

این است که pdf توزیع قابل مشتق باشد.

مثال:

اگر $\theta = \lambda$

$X \sim pos(\lambda)$ $\quad$ UMVUE = ?

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} I_{\{0,1,\dots\}}(x)$$

$$\ln f(x;\lambda) = x\ln\lambda - \lambda - \ln x! + \ln I_{\{0,1,\dots\}}(x)$$

$$\frac{\partial \ln f_X(x;\lambda)}{\partial\lambda} = \frac{x}{\lambda} - 1$$

$x \to X$

$$\frac{X}{\lambda} - 1 \qquad E\left(\left(\frac{X}{\lambda}-1\right)^2\right) = \frac{1}{\lambda^2}E\left((X-\lambda)^2\right) = \frac{1}{\lambda}$$

$$CR(n,\theta) = \frac{1}{n} = \frac{\lambda}{n}$$

حال اگر یک برآورد دیگر پیدا کنیم که واریانسش برابر با $\frac{\lambda}{n}$ باشد UMVUE است.

برای مثال $T = \bar{X}$

$$E(\bar{X}) = \mu \qquad var(\bar{X}) = \frac{\sigma^2}{n} = \frac{\lambda}{n}$$

پس UMVUE : $\bar{X}$