(https://stanford.edu/%7Eshervine/teaching/cs-229.html)

یادآوری آمار و احتمالات

مقدمه‌ای بر احتمالات و ترکیبیات

فضای نمونه - مجموعه‌ی همه‌ی پیشامدهای یک آزمایش را فضای نمونه‌ی آن آزمایش گویند که با S نمایش داده می‌شود.

رخداد — هر زیرمجموعه‌ی E از فضای نمونه یک رخداد در نظر گرفته می‌شود. به عبارت دیگر، یک رخداد مجموعه‌ای از پیشامدهای یک آزمایش است. اگر پیشامد یک آزمایش عضوی از مجموعه‌ی E باشد، در این حالت می‌گوییم که رخداد E اتفاق افتاده است.

اصول موضوعه‌ی احتمالات — برای هر رخداد E، $P(E)$ احتمال اتفاق افتادن رخداد E می‌باشد.

اصل ۱ - احتمال عددی بین ۰ و ۱ است.

$$\boxed{0 \leqslant P(E) \leqslant 1}$$

اصل ۲ - احتمال اینکه حداقل یکی از رخدادهای موجود در فضای نمونه اتفاق بیوفتد، ۱ است.

$$\boxed{P(S) = 1}$$

اصل ۳ - برای هر دنباله از رخدادهایی که دو به دو اشتراک نداشته باشند، داریم:

$$\boxed{P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)}$$

جایگشت - یک جایگشت چیدمانی از r شی از n شی با یک ترتیب خاص است. تعداد این چنین جایگشت‌ها $P(n,r)$ است که به صورت زیر تعریف می‌شود:

$$\boxed{P(n,r) = \frac{n!}{(n-r)!}}$$

ترکیب - یک ترکیب چیدمانی از r شی از n شی است، به طوری که ترتیب اهمیتی نداشته باشد. تعداد این چنین ترکیب‌ها $C(n,r)$ است که به صورت زیر تعریف می‌شود:

$$\boxed{C(n,r) = \frac{P(n,r)}{r!} = \frac{n!}{r!(n-r)!}}$$

نکته: برای $0 \leqslant r \leqslant n$*، داریم* $P(n,r) \geqslant C(n,r)$

احتمال شرطی

قضیه‌ی بیز - برای رخدادهای A و B به طوری که $P(B) > 0$ داریم:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

نکته: داریم $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$

افراز - فرض می‌کنیم برای $\{A_i, i \in [\![1, n]\!]\}$ به ازای هر i داشته باشیم $A_i \neq \varnothing$. در این صورت می‌گوییم $\{A_i\}$ یک افراز است اگر:

$$\boxed{\forall i \neq j, A_i \cap A_j = \emptyset \quad \text{et} \quad \bigcup_{i=1}^{n} A_i = S}$$

نکته: برای هر رخداد B در فضای نمونه داریم $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$.

تعمیم قضیه‌ی بیز - فرض می‌کنیم $\{A_i, i \in [\![1, n]\!]\}$ یک افراز از فضای نمونه باشید. در این صورت داریم:

$$\boxed{P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}}$$

استقلال - دو رخداد A و B مستقل هستند اگر و فقط اگر داشته باشیم:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

متغیرهای تصادفی

تعاریف

متغیر تصادفی - یک متغیر تصادفی، که معمولاً با X نمایش داده می‌شود، یک تابع است که هر عضو فضای نمونه را به اعداد حقیقی نگاشت می‌کند.

تابع توزیع تجمعی - تابع توزیع تجمعی F، که تابعی یکنوا و اکیدا غیرنزولی است و برای آن $\lim\limits_{x \to -\infty} F(x) = 0$ و $\lim\limits_{x \to +\infty} F(x) = 1$ صدق می‌کنید، به صورت زیر تعریف می‌شود:

$$\boxed{F(x) = P(X \leqslant x)}$$

نکته: داریم $P(a < X \leqslant B) = F(b) - F(a)$.

تابع چگالی احتمال **(PDF)** - تابع چگالی احتمال f احتمال آن است که متغیر تصادفی X مقداری بین دو تحقق همجوار این متغیر تصادفی را بگیرد.

ارتباط بین **PDF** و **CDF** - موارد زیر ویژگی‌های مهمی هستند که باید در مورد حالت گسسته و حالت پیوسته در نظر گرفت.

حالت	**CDF** F	**PDF** f	**ویژگی‌های PDF**
(D)	$F(x) = \sum\limits_{x_i \leqslant x} P(X = x_i)$	$f(x_j) = P(X = x_j)$	$0 \leqslant f(x_j) \leqslant 1 \text{ et } \sum\limits_{j} f(x_j) = 1$
(C)	$F(x) = \int_{-\infty}^{x} f(y)dy$	$f(x) = \frac{dF}{dx}$	$f(x) \geqslant 0 \text{ et } \int_{-\infty}^{+\infty} f(x)dx = 1$

امید ریاضی و گشتاورهای یک توزیع - عبارت‌های مربوط به امید ریاضی $E[X]$، امید ریاضی تعمیم یافته $E[g(X)]$، k-مین گشتاور $E[X^k]$، و تابع ویژگی $\psi(\omega)$ برای حالات پیوسته و گسسته به صورت زیر هستند:

حالت	$E[X]$	$E[g(X)]$	$E[X^k]$	$\psi(\omega)$

(D)	$\sum_{i=1}^{n} x_i f(x_i)$	$\sum_{i=1}^{n} g(x_i) f(x_i)$	$\sum_{i=1}^{n} x_i^k f(x_i)$	$\sum_{i=1}^{n} f(x_i) e^{i\omega x_i}$
(C)	$\int_{-\infty}^{+\infty} x f(x) dx$	$\int_{-\infty}^{+\infty} g(x) f(x) dx$	$\int_{-\infty}^{+\infty} x^k f(x) dx$	$\int_{-\infty}^{+\infty} f(x) e^{i\omega x} dx$

واریانس - واریانس یک متغیر تصادفی، که معمولاً با $\mathrm{Var}(X)$ یا σ^2 نمایش داده می‌شود، میزانی از پراکندگی یک تابع توزیع است. مقدار واریانس به صورت زیر به دست می‌آید:

$$\boxed{\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2}$$

انحراف معیار - انحراف معیار یک متغیر تصادفی، که با σ نمایش داده می‌شود، میزانی از پراکندگی یک تابع توزیع است که با متغیر تصادفی هم‌واحد است. مقدار آن به صورت زیر به دست می‌آید:

$$\boxed{\sigma = \sqrt{\mathrm{Var}(X)}}$$

تبدیلات متغیرهای تصادفی - فرض کنید متغیرهای تصادفی X و Y توسط تابعی به هم مرتبط هستند. با نمایش تابع توزیع متغیرهای تصادفی X و Y با f_X و f_Y داریم:

$$\boxed{f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|}$$

قضیه‌ی انتگرال لایبنیتس - فرض کنید g تابعی از x و c باشد، و a و b کران‌هایی باشند که مقدار آن‌ها وابسته به مقدار c باشد. داریم:

$$\boxed{\frac{\partial}{\partial c}\left(\int_a^b g(x)dx\right) = \frac{\partial b}{\partial c}\cdot g(b) - \frac{\partial a}{\partial c}\cdot g(a) + \int_a^b \frac{\partial g}{\partial c}(x)dx}$$

توزیع‌های احتمالی

نابرابری چبیشف - فرض کنید X متغیری تصادفی با امید ریاضی μ باشد. برای هر k و $\sigma > 0$ نابرابری زیر را داریم:

$$\boxed{P(|X-\mu| \geqslant k\sigma) \leqslant \frac{1}{k^2}}$$

توزیع‌های احتمالی اصلی - توزیع‌های زیر توزیع‌های احتمالی اصلی هستند که بهتر است به خاطر بسپارید:

نوع	توزیع	**PDF**	$\psi(\omega)$	$E[X]$	$\mathrm{Var}(X)$
(D)	$X \sim \mathcal{B}(n,p)$	$P(X=x) = \binom{n}{x} p^x q^{n-x}$	$(pe^{i\omega}+q)^n$	np	npq
(D)	$X \sim \mathrm{Po}(\mu)$	$P(X=x) = \frac{\mu^x}{x!} e^{-\mu}$	$e^{\mu(e^{i\omega}-1)}$	μ	μ
(C)	$X \sim \mathcal{U}(a,b)$	$f(x) = \frac{1}{b-a}$	$\frac{e^{i\omega b}-e^{i\omega a}}{(b-a)i\omega}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
(C)	$X \sim \mathcal{N}(\mu,\sigma)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$	μ	σ^2
(C)	$X \sim \mathrm{Exp}(\lambda)$	$f(x) = \lambda e^{-\lambda x}$	$\frac{1}{1-\frac{i\omega}{\lambda}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

متغیرهای تصادفی با توزیع مشترک

چگالی حاشیه‌ای و توزیع تجمعی - از تابع چگالی احتمالی مشترک f_{XY} داریم:

حالت	چگالی حاشیه‌ای	تابع تجمعی

حالت	چگالی حاشیه‌ای	تابع تجمعی
(D)	$f_X(x_i) = \sum_j f_{XY}(x_i, y_j)$	$F_{XY}(x, y) = \sum_{x_i \leqslant x} \sum_{y_j \leqslant y} f_{XY}(x_i, y_j)$
(C)	$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$	$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(x', y') dx' dy'$

چگالی شرطی - چگالی شرطی X نسبت به Y، که معمولاً با $f_{X|Y}$ نمایش داده می‌شود، به صورت زیر تعریف می‌شود:

$$\boxed{f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}}$$

استقلال - دو متغیر تصادفی X و Y مستقل هستند اگر داشته باشیم:

$$\boxed{f_{XY}(x, y) = f_X(x) f_Y(y)}$$

کواریانس - کواریانس دو متغیر تصادفی X و Y که با σ_{XY}^2 یا به صورت معمول‌تر با $\mathrm{Cov}(X, Y)$ نمایش داده می‌شود، به صورت زیر است:

$$\boxed{\mathrm{Cov}(X, Y) \triangleq \sigma_{XY}^2 = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y}$$

همبستگی - با نمایش انحراف معیار X و Y به صورت σ_X و σ_Y، همبستگی مابین دو متغیر تصادفی X و Y که با ρ_{XY} نمایش داده می‌شود به صورت زیر تعریف می‌شود:

$$\boxed{\rho_{XY} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}}$$

نکته‌ی ۱: برای هر دو متغیر تصادفی دلخواه X و Y، داریم $\rho_{XY} \in [-1, 1]$.

نکته‌ی ۲: اگر X و Y مستقل باشند، داریم $\rho_{XY} = 0$.

تخمین پارامتر

تعاریف

نمونه‌ی تصادفی - یک نمونه‌ی تصادفی مجموعه‌ای از n متغیر تصادفی $X_1, \ldots, X_n$ است که از هم مستقل هستند و توزیع یکسانی با X دارند.

تخمین‌گر - یک تخمین‌گر تابعی از داده‌ها است که برای به‌دست‌آوردن مقدار نامشخص یک پارامتر در یک مدل آماری به کار می‌رود.

پیش‌قدر - پیش‌قدر یک تخمین‌گر $\hat{\theta}$ به عنوان اختلاف بین امید ریاضی توزیع $\hat{\theta}$ و مقدار واقعی تعریف می‌شود. یعنی:

$$\boxed{\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta}$$

نکته: یک تخمین‌گر بدون پیش‌قدر است اگر داشته باشیم $E[\hat{\theta}] = \theta$.

تخمین میانگین

میانگین نمونه - میانگین نمونه‌ی یک نمونه‌ی تصادفی که برای تخمین مقدار واقعی میانگین μ یک توزیع به کار می‌رود، معمولاً با $\overline{X}$ نمایش داده می‌شود و به صورت زیر تعریف می‌شود:

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

نکته: میانگین نمونه بدون پیش‌قدر است، یعنی $E[\overline{X}] = \mu$.

قضیه‌ی حد مرکزی - یک نمونه‌ی تصادفی $X_1, ..., X_n$ که از یک توزیع با میانگین μ و واریانس σ^2 به دست آمده‌اند را در نظر بگیرید؛ داریم:

$$\boxed{\overline{X} \underset{n\to+\infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

تخمین واریانس

واریانس نمونه - واریانس نمونه‌ی یک نمونه‌ی تصادفی که برای تخمین مقدار واقعی واریانس σ^2 یک توزیع به کار می‌رود، معمولاً با s^2 یا $\hat{\sigma}^2$ نمایش داده می‌شود و به صورت زیر تعریف می‌شود:

$$\boxed{s^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

نکته: واریانس نمونه بدون پیش‌قدر است، یعنی $E[s^2] = \sigma^2$.

رابطه‌ی χ^2 با واریانس نمونه - فرض کنید s^2 واریانس نمونه‌ی یک نمونه‌ی تصادفی باشد. داریم:

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