- $\text{Cov}[aX, bY] = ab\text{Cov}[X, Y]$
- $\text{Cov}[X + a, Y + b] = \text{Cov}[X, Y]$
- $\text{Cov}\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}[X_i, Y_j]$

Correlation

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\,\mathbb{V}[Y]}}$$

Independence

$$X \perp\!\!\!\perp Y \implies \rho[X, Y] = 0 \iff \text{Cov}[X, Y] = 0 \iff \mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$$

Sample variance

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

Conditional variance

- $\mathbb{V}[Y \mid X] = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X])^2 \mid X\right] = \mathbb{E}\left[Y^2 \mid X\right] - \mathbb{E}[Y \mid X]^2$
- $\mathbb{V}[Y] = \mathbb{E}[\mathbb{V}[Y \mid X]] + \mathbb{V}[\mathbb{E}[Y \mid X]]$

6 Inequalities

Cauchy-Schwarz

$$\mathbb{E}[XY]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

Markov

$$\mathbb{P}[\varphi(X) \geq t] \leq \frac{\mathbb{E}[\varphi(X)]}{t}$$

Chebyshev

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \leq \frac{\mathbb{V}[X]}{t^2}$$

Chernoff

$$\mathbb{P}[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right) \quad \delta > -1$$

Hoeffding

$$X_1, \ldots, X_n \text{ independent } \wedge\ \mathbb{P}[X_i \in [a_i, b_i]] = 1\ \wedge\ 1 \leq i \leq n$$

$$\mathbb{P}\left[\bar{X} - \mathbb{E}\left[\bar{X}\right] \geq t\right] \leq e^{-2nt^2} \quad t > 0$$

$$\mathbb{P}\left[|\bar{X} - \mathbb{E}\left[\bar{X}\right]| \geq t\right] \leq 2\exp\left\{-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right\} \quad t > 0$$

Jensen

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]) \quad \varphi \text{ convex}$$

7 Distribution Relationships

Binomial

- $X_i \sim \text{Bern}(p) \implies \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$
- $X \sim \text{Bin}(n, p), Y \sim \text{Bin}(m, p) \implies X + Y \sim \text{Bin}(n + m, p)$
- $\lim_{n\to\infty} \text{Bin}(n, p) = \text{Po}(np) \qquad (n \text{ large, } p \text{ small})$
- $\lim_{n\to\infty} \text{Bin}(n, p) = \mathcal{N}(np, np(1-p)) \qquad (n \text{ large, } p \text{ far from 0 and 1})$

Negative Binomial

- $X \sim \text{NBin}(1, p) = \text{Geo}(p)$
- $X \sim \text{NBin}(r, p) = \sum_{i=1}^{r} \text{Geo}(p)$
- $X_i \sim \text{NBin}(r_i, p) \implies \sum X_i \sim \text{NBin}\left(\sum r_i, p\right)$
- $X \sim \text{NBin}(r, p).\ Y \sim \text{Bin}(s + r, p) \implies \mathbb{P}[X \leq s] = \mathbb{P}[Y \geq r]$

Poisson

- $X_i \sim \text{Po}(\lambda_i) \wedge X_i \perp\!\!\!\perp X_j \implies \sum_{i=1}^{n} X_i \sim \text{Po}\left(\sum_{i=1}^{n} \lambda_i\right)$
- $X_i \sim \text{Po}(\lambda_i) \wedge X_i \perp\!\!\!\perp X_j \implies X_i \left| \sum_{j=1}^{n} X_j \right. \sim \text{Bin}\left(\sum_{j=1}^{n} X_j, \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}\right)$

Exponential

- $X_i \sim \text{Exp}(\beta) \wedge X_i \perp\!\!\!\perp X_j \implies \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \beta)$
- Memoryless property: $\mathbb{P}[X > x + y \mid X > y] = \mathbb{P}[X > x]$

Normal

- $X \sim \mathcal{N}\left(\mu, \sigma^2\right) \implies \left(\frac{X-\mu}{\sigma}\right) \sim \mathcal{N}(0, 1)$
- $X \sim \mathcal{N}\left(\mu, \sigma^2\right) \wedge Z = aX + b \implies Z \sim \mathcal{N}\left(a\mu + b, a^2\sigma^2\right)$
- $X_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right) \wedge X_i \perp\!\!\!\perp X_j \implies \sum_i X_i \sim \mathcal{N}\left(\sum_i \mu_i, \sum_i \sigma_i^2\right)$
- $\mathbb{P}[a < X \leq b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$
- $\Phi(-x) = 1 - \Phi(x) \qquad \phi'(x) = -x\phi(x) \qquad \phi''(x) = (x^2 - 1)\phi(x)$
- Upper quantile of $\mathcal{N}(0, 1)$: $z_\alpha = \Phi^{-1}(1 - \alpha)$

Gamma

- $X \sim \text{Gamma}(\alpha, \beta) \iff X/\beta \sim \text{Gamma}(\alpha, 1)$
- $\text{Gamma}(\alpha, \beta) \sim \sum_{i=1}^{\alpha} \text{Exp}(\beta)$
- $X_i \sim \text{Gamma}(\alpha_i, \beta) \wedge X_i \perp\!\!\!\perp X_j \implies \sum_i X_i \sim \text{Gamma}\left(\sum_i \alpha_i, \beta\right)$

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

Beta

- $\frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
- $\mathbb{E}\left[X^k\right] = \frac{\mathrm{B}(\alpha+k,\beta)}{\mathrm{B}(\alpha,\beta)} = \frac{\alpha+k-1}{\alpha+\beta+k-1}\mathbb{E}\left[X^{k-1}\right]$
- $\mathrm{Beta}\,(1,1) \sim \mathrm{Unif}\,(0,1)$

8 Probability and Moment Generating Functions

- $G_X(t) = \mathbb{E}\left[t^X\right] \qquad |t| < 1$
- $M_X(t) = G_X(e^t) = \mathbb{E}\left[e^{Xt}\right] = \mathbb{E}\left[\sum_{i=0}^\infty \frac{(Xt)^i}{i!}\right] = \sum_{i=0}^\infty \frac{\mathbb{E}\left[X^i\right]}{i!} \cdot t^i$
- $\mathbb{P}\left[X=0\right] = G_X(0)$
- $\mathbb{P}\left[X=1\right] = G'_X(0)$
- $\mathbb{P}\left[X=i\right] = \frac{G_X^{(i)}(0)}{i!}$
- $\mathbb{E}\left[X\right] = G'_X(1^-)$
- $\mathbb{E}\left[X^k\right] = M_X^{(k)}(0)$
- $\mathbb{E}\left[\frac{X!}{(X-k)!}\right] = G_X^{(k)}(1^-)$
- $\mathbb{V}\left[X\right] = G''_X(1^-) + G'_X(1^-) - \left(G'_X(1^-)\right)^2$
- $G_X(t) = G_Y(t) \implies X \stackrel{d}{=} Y$

9 Multivariate Distributions

9.1 Standard Bivariate Normal

Let $X, Y \sim \mathcal{N}(0,1) \wedge X \perp\!\!\!\perp Z$ where $Y = \rho X + \sqrt{1-\rho^2} Z$

Joint density

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right\}$$

Conditionals

$$(Y \mid X = x) \sim \mathcal{N}\left(\rho x, 1-\rho^2\right) \qquad \text{and} \qquad (X \mid Y = y) \sim \mathcal{N}\left(\rho y, 1-\rho^2\right)$$

Independence

$$X \perp\!\!\!\perp Y \iff \rho = 0$$

9.2 Bivariate Normal

Let $X \sim \mathcal{N}\left(\mu_x, \sigma_x^2\right)$ and $Y \sim \mathcal{N}\left(\mu_y, \sigma_y^2\right)$.

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{z}{2(1-\rho^2)}\right\}$$

$$z = \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]$$

Conditional mean and variance

$$\mathbb{E}\left[X \mid Y\right] = \mathbb{E}\left[X\right] + \rho\frac{\sigma_X}{\sigma_Y}(Y - \mathbb{E}\left[Y\right])$$

$$\mathbb{V}\left[X \mid Y\right] = \sigma_X\sqrt{1-\rho^2}$$

9.3 Multivariate Normal

Covariance matrix Σ (Precision matrix Σ^{-1})

$$\Sigma = \begin{pmatrix} \mathbb{V}\left[X_1\right] & \cdots & \mathrm{Cov}\left[X_1, X_k\right] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left[X_k, X_1\right] & \cdots & \mathbb{V}\left[X_k\right] \end{pmatrix}$$

If $X \sim \mathcal{N}(\mu, \Sigma)$,

$$f_X(x) = (2\pi)^{-n/2}\left|\Sigma\right|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Properties

- $Z \sim \mathcal{N}(0,1) \wedge X = \mu + \Sigma^{1/2}Z \implies X \sim \mathcal{N}(\mu, \Sigma)$
- $X \sim \mathcal{N}(\mu, \Sigma) \implies \Sigma^{-1/2}(X-\mu) \sim \mathcal{N}(0,1)$
- $X \sim \mathcal{N}(\mu, \Sigma) \implies AX \sim \mathcal{N}\left(A\mu, A\Sigma A^T\right)$
- $X \sim \mathcal{N}(\mu, \Sigma) \wedge \|a\| = k \implies a^T X \sim \mathcal{N}\left(a^T\mu, a^T\Sigma a\right)$

10 Convergence

Let $\{X_1, X_2, \ldots\}$ be a sequence of RV's and let X be another RV. Let F_n denote the CDF of X_n and let F denote the CDF of X.
Types of Convergence

1. In distribution (weakly, in law): $X_n \xrightarrow{\text{D}} X$

$$\lim_{n\to\infty} F_n(t) = F(t) \qquad \forall t \text{ where } F \text{ continuous}$$