Statistic

For robust statistics, see the page.

Textbook

Matrix cookbook

- The matrix cookbook
- a for scalar, $\underline{a}$ for vector, A for matrix.
- X for random variable, $\underline{X}$ for random vector, $\mathbf{X}$ for random matrix.
- $\mathbb{E}[\underline{a}^T\underline{X}] = \underline{a} \cdot \mathbb{E}\underline{X}$; $\mathbb{E}[A\underline{X}] = A\mathbb{E}\underline{X}$; $\mathbb{E}A\mathbf{X}B = A\mathbb{E}[\mathbf{X}]B$.
- $\underline{\mu} := \mathbb{E}\underline{X}, \underline{v} := \mathbb{E}\underline{Y}$.

Covariance

1. $Cov[X, Y] = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$.
2. $V[\frac{X}{\sigma_X}] = 1$, $\rho_{XY} = Cov[\tilde{X}, \tilde{Y}] = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$
3. The **covariance matrix** $V[\underline{X}] = \mathbb{E}[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})^T] = \mathbb{E}\underline{X}\underline{X}^T - \underline{\mu}\underline{\mu}^T$.
4. The **variance of r.v.** in the form of $\underline{a} \cdot \underline{X}$: $V[\underline{a} \cdot \underline{X}] = \underline{a}^T V[\underline{X}]\underline{a}$; $V[A\underline{X}] = AV[\underline{X}]A^T$. When $||\underline{a}|| = 1$, it's taking the variance in the direction of $\underline{a}$, over the joint p.d.f. (consider variance as the width of the function picture along certain axis).

Multivariate normal

1. Standard normal: $\underline{Z} \sim N(\underline{0}, I)$.
2. (Stretch) $\underline{X} = D\underline{Z}$ where $D = diag(\sigma_1, \sigma_2, \ldots)$, then $\underline{X} \sim N(\underline{0}, D^2)$.
3. (Rotate) $\underline{Y} = Q\underline{X}$ where $Q^TQ = I$ (orthogonal), then $\underline{Y} \sim N(\underline{0}, V)$. Given D to find V and Q, consider that by (Covariance-4),

$$V = V[Y] = V[QX] = QV[X]Q^T = QD^2Q^T$$

> Eigenvalue decomposition solves this problem. Say $eigen(V)=\lambda_{1}, \dots$, then $\lambda_{i}=\sigma_{i}^{2}$.

Hypothesis Testing

Kullback–Leibler divergence

a measure of how one probability distribution P is different from a second, reference probability distribution Q.

STAT 400

> Def. (**Bernoulli distribution**, $X \sim Be(p)$) Bernoulli trial is that $A \in \mathcal{F}$, and we call the trial a success if A occurs. Bernoulli distribution is based on single Bernoulli trial.

> Def. (**Binomial distribution**, $Y \sim Bin(n, p)$) Perform n independent Bernoulli trials with $p = \mathbb{P}(A)$, and let Bernoulli r.v.s. $X_1, X_2 \dots X_n$ be the indicator funciton of success of the experiments. Let $Y := \sum X_i$, then the p.m.f. $f_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 1, 2, \dots, n$. **Binomial process** is that Y_n to be the number of successes in the first n $Be(p)$ trials.

> Def. (**Geometric distribution**, $W \sim Geom(p)$) Keep performing Bernoulli trials until the first success, and let $W := \text{the waiting time}$, then $f_W(k) = (1-p)^{k-1} p, k = 1, 2, \dots, F(x) = 1 - (1-p)^x$.

> Def. (**Negative Binomial distribution**, $W_r \sim NB(r, p)$) Let $W_r := \text{the waiting time for r successes}$, then $f_{W_r}(k) = \binom{k-1}{k-r} p^{r-1} (1-p)^{k-r} p, k = 1, 2, \dots$

> Def. (**Poisson distribution**, $X \sim poisson(\lambda)$) $f_X(k) := \frac{\lambda^k}{k!} e^{-\lambda}, k = 1, 2, \dots$ The Poisson distribution is an approximation of a binomial distribution of a rare event in the case that **n is large, p is small**, and $\lambda = np$ is moderate.

> Def. (**Hypergeometric distribution**) $f_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$. $\mathbb{E}X = nN_1/N, VarX = n\frac{N_1}{N}\frac{N-K}{N}\frac{N-n}{N-1}$. The sample

> Def. (**Multinomial**) Multi-outcome version of hypergeometric. $f_X(x) = \frac{n!}{\prod x_i!} \prod p_i^{x_i}$, where $\mathbb{E}X_i = np_i, VarX_i = np_i(1-p_i)$.