

Statistics

The note of UIUC course STAT 400 in Fall 2023, by Zory Zhang. In case of any broken math rendering in Github preview, please open this markdown file using your own markdown editor. PDF version (maybe not up-to-date) can be found [here](#).

For robust statistics, see [the page](#).

Preliminaries

The following content is based on the course STAT400@UIUC. [Probability knowledge](#) is assumed.

Textbook

Ch0 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.
- $\mathbb{E}[\underline{a}^T \underline{X}] = \underline{a} \cdot \mathbb{E}\underline{X}$; $\mathbb{E}[A\underline{X}] = A\mathbb{E}\underline{X}$; $\mathbb{E}A\mathbf{X}B = A\mathbb{E}[\mathbf{X}]B$.
- $\underline{\mu} := \mathbb{E}\underline{X}$, $\underline{v} := \mathbb{E}\underline{Y}$.

Ch1 Distributions

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 function 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 $W :=$ 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 $W_r :=$ the waiting time for r successes, then $f_{W_r}(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$, $k = 1, 2, \dots$.

Def. (**Approximate poisson process** with param $\lambda > 0$) Let the number of **occurrences** of some event in a given **continuous interval** be counted. If

1. The number of occurrences in nonoverlapping subintervals are independent.
2. The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh .
3. The probability of more than one occurrence in a sufficiently short subinterval of length h is essentially 0.

Def. (**Poisson distribution**, $X \sim poisson(\lambda)$) An approximation of number of occurrences in an interval of length 1 is given. Consider partition the unit interval into n subintervals. As n becomes larger, the length of subinterval become smaller and thus more likely the last two properties hold. For a large n , $\mathbb{P}(X = k) \sim \binom{n}{k} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k}$, and it converges to $f_X(k) := \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$

Rmk. $M(t) = e^{\lambda(e^t-1)}$, then $\mu = \sigma^2 = \lambda$.

E.g. Number of customers of a shop between 5-6pm.

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 $\mathbb{E}X_i = np_i$, $VarX_i = np_i(1 - p_i)$.

Def. (**Exponential distribution**, $X \sim Exp(\lambda)$) Define X as the waiting time for the first occurrence in a poisson process. Then $F_X(x) = 1 - \mathbb{P}(\text{no occurrences in } [0, x])$, therefore

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}, f_X(x) = F'_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

Rmk.

1. It's usually expressed as $\theta = \frac{1}{\lambda}$, $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$.
2. $M(t) = \frac{1}{1-\theta t}$ for $t < \frac{1}{\theta}$, thus $\mu = \theta$, $\sigma^2 = \theta^2$, Fisher's moment coefficient of skewness is 2.
3. (Memoryless) $\mathbb{P}(X > s+t | X > t) = \mathbb{P}(X > s)$.
4. (Failure/hazard rate for positive r.v.) $r(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}(X \leq t+h | X > t) = \frac{f(t)}{1-F(t)} = \lambda$.

Def. (**Gamma function**) For $t > 0$, $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$.

Rmk. $\Gamma(t) = (-y^{t-1} e^{-y})_{y=0}^\infty + \int_0^\infty (t-1)y^{t-2} e^{-y} dy = (t-1)\Gamma(t-1)$. $\Gamma(1) = 1$. When t is an integer, $\Gamma(t) = (t-1)!$.

Def. (**Gamma distribution**, $X \sim Gamma(\alpha, \theta)$) For $x \geq 0$, $f_X(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$. In a poisson process, the waiting time of first occurrence is $Exp(\theta = 1/\lambda)$, now gamma distribution describes the waiting time until α -th occurrence.

Rmk. $\mathbb{E}X = \alpha\theta$, $VarX = \alpha\theta^2$.

Def. (**Normal/Gaussian distribution**, $X \sim \mathcal{N}(\mu, \sigma^2)$) $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$. Standard normal distribution $Y \sim \mathcal{N}(0, 1)$, denote the p.d.f. as $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.

Rmk. To compute $I = \int_{-\infty}^\infty \phi(x) dx$, consider

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \phi(x) dx \int_{-\infty}^{\infty} \phi(y) dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy \\
&= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{r^2}{2}\right) r dr d\theta = 1
\end{aligned}$$

Rmk. For normal distribution, the i -th central moment of $N(\mu, \sigma^2)$ is $\sigma_i = \begin{cases} 0, & \text{if } i \text{ odd,} \\ \sigma^i (i-1)!!, & \text{o.w.} \end{cases}$

Def. ((**standard**) **Cauchy distribution**, $X \sim \mathcal{N}(\mu, \sigma^2)$) p.d.f.

$$f_X(x) = \frac{1}{\pi(1+x^2)}, F_X(x) = \arctan(x) + \frac{\pi}{2}.$$

Rmk. (Heavy tail distribution without moment) Notice the $\mathbb{E}X = \int_{-\infty}^{\infty} |x|f_X(x) dx = \infty$ does not exist.

Ch2

2.1 Covariance

1. $\sigma_{XY} = \text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}(XY) - \mu_X \mu_Y$.
2. $V[\frac{X}{\sigma_X}] = 1, \rho_{XY} = \text{Cov}[\tilde{X}, \tilde{Y}] = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} \in [-1, 1]$.
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).

2.2

Prop. (**Independent normal**) $\sum_{i=1}^n c_i X_i \sim N(\sum c_i \mu_i, \sum c_i^2 \sigma_i^2)$.

Cor. When all X_i are from the same distribution, $\frac{1}{n} \sum X_i \sim N(\mu, \frac{\sigma^2}{n})$.

Thm. (**CLT**) $\frac{1}{n} S_n := \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$.

Multivariate normal

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

$$V = V[\underline{Y}] = V[Q\underline{X}] = QV[\underline{X}]Q^T = QD^2Q^T$$

Eigenvalue decomposition solves this problem. Say $\text{eigen}(V) = \lambda_1, \dots$, 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 .