Formula Sheet – Introduction to Data Science (1MS041)

Based on Lecture Notes and All of Statistics – Wasserman (Ch. 1–5, 9)

---

1. Probability Basics

$$
\begin{aligned}
P(A^c) &= 1 - P(A) \\
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
P(A|B) &= \frac{P(A \cap B)}{P(B)} \quad (P(B) > 0) \\
P(A_i|B) &= \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \quad \text{(Bayes' Theorem)} \\
P(A \cap B) &= P(A)P(B) \quad \text{if A, B independent}
\end{aligned}
$$

2. Random Variables and Expectation

$$
\begin{aligned}
E[X] &= \begin{cases} \sum_x x f(x), & \text{discrete} \\ \int_{-\infty}^{\infty} x f(x)\,dx, & \text{continuous} \end{cases} \\
Var(X) &= E[(X - E[X])^2] = E[X^2] - (E[X])^2 \\
Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
Z &= \frac{X - \mu}{\sigma} \quad \text{(Standardization)} \\
E[aX + bY] &= aE[X] + bE[Y] \\
E[XY] &= E[X]E[Y] \quad \text{if X,Y independent}
\end{aligned}
$$

3. Variance and Standard Deviation

$$\sigma = \sqrt{Var(X)} = \sqrt{E[(X - E[X])^2]}$$

For sample data $x_1, \ldots, x_n$:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

4. Common Probability Distributions

Discrete

Distribution	PMF	Mean / Variance
Bernoulli(p)	$f(x) = p^x(1-p)^{1-x},\ x \in \{0,1\}$	$E[X] = p,\ Var[X] = p(1-p)$
Binomial(n, p)	$f(x) = \binom{n}{x}p^x(1-p)^{n-x}$	$E[X] = np,\ Var[X] = np(1-p)$
Geometric(p)	$f(x) = p(1-p)^{x-1},\ x = 1, 2, \ldots$	$E[X] = \frac{1}{p},\ Var[X] = \frac{1-p}{p^2}$
Poisson(λ)	$f(x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2, \ldots$	$E[X] = Var[X] = \lambda$
Discrete Uniform(k)	$f(x) = \frac{1}{k},\ x = 1, \ldots, k$	$E[X] = \frac{k+1}{2},\ Var[X] = \frac{k^2-1}{12}$

Continuous

Distribution	PDF	Mean / Variance
Uniform(a, b)	$f(x) = \frac{1}{b-a},\ a \le x \le b$	$E[X] = \frac{a+b}{2},\ Var[X] = \frac{(b-a)^2}{12}$
Exponential(λ)	$f(x) = \lambda e^{-\lambda x},\ x \ge 0$	$E[X] = \frac{1}{\lambda},\ Var[X] = \frac{1}{\lambda^2}$
Normal(μ, σ^2)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$E[X] = \mu,\ Var[X] = \sigma^2$
Gamma(α, β)	$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}$	$E[X] = \alpha\beta,\ Var[X] = \alpha\beta^2$
Beta(α, β)	$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$E[X] = \frac{\alpha}{\alpha+\beta}$

5. Distribution Functions

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

For discrete variables:

$$F(x) = \sum_{t \le x} f(t)$$

6. Important Inequalities

$$P(X \ge a) \le \frac{E[X]}{a} \quad \text{(Markov)}$$
$$P(|X - E[X]| \ge \epsilon) \le \frac{Var(X)}{\epsilon^2} \quad \text{(Chebyshev)}$$
$$P(|\bar{X} - E[X]| \ge \epsilon) \le 2e^{-\frac{2n\epsilon^2}{(b-a)^2}} \quad \text{(Hoeffding)}$$

7. Limit Theorems

Law of Large Numbers (LLN)

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E[X]$$

Central Limit Theorem (CLT)

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance $\sigma^2 < \infty$. Then

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

or equivalently,

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

Expectation and Variance of the Sample Mean

Property	**Formula**
Expectation of sample mean	$E[\bar{X}_n] = E\big[\frac{1}{n}\sum_{i=1}^n X_i\big] = \mu$
Variance of sample mean	$Var(\bar{X}_n) = Var\big(\frac{1}{n}\sum_{i=1}^n X_i\big) = \frac{\sigma^2}{n}$
Standard deviation (standard error)	$SD(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$

8. Risk and Learning Theory

$$R(g) = E[L(Z,g)] = \int L(z,g)\,dF(z)$$
$$R_n(g) = \frac{1}{n}\sum_{i=1}^n L(Z_i, g)$$
$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$$

Bias

$$Bias(\hat{\theta};\theta) = \mathbb{E}_\theta[\hat{\theta}] - \theta$$
$$\text{Unbiasedness:} \quad \mathbb{E}_\theta[\hat{\theta}] = \theta$$
$$\text{Relative Bias:} \quad \frac{\mathbb{E}_\theta[\hat{\theta}] - \theta}{\theta} \quad (\theta \neq 0)$$

Pointwise Bias for Function Estimation

For a regression or function estimation target f at input x:

$$Bias(\hat{f}(x)) = \mathbb{E}[\hat{f}(x)] - f(x)$$
$$\mathbb{E}\big[(\hat{f}(x) - f(x))^2\big] = Var(\hat{f}(x)) + \big(Bias(\hat{f}(x))\big)^2$$

9. Norms and Convergence

$$\|X\|_p = (E[|X|^p])^{1/p}$$

$$X_n \xrightarrow{P} X \iff P(|X_n - X| > \epsilon) \to 0$$
$$X_n \xrightarrow{a.s.} X \iff P(\lim X_n = X) = 1$$
$$X_n \xrightarrow{d} X \iff F_{X_n}(x) \to F_X(x)$$
$$X_n \xrightarrow{L^2} X \iff E[(X_n - X)^2] \to 0$$

10. Likelihood and Maximum Likelihood Estimation (MLE)

Likelihood Function

Given independent observations $x_1, x_2, \ldots, x_n$ from a model with pdf/pmf $f(x;\theta)$:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Log-Likelihood Function

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

Maximum Likelihood Estimator (MLE)

The MLE is the parameter value that maximizes the likelihood (or log-likelihood):

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \ell(\theta)$$

Common MLEs

Distribution	**Parameter**	**MLE**
Bernoulli	θ	$\hat{\theta} = \bar{X}$
Binomial(n, θ)	θ	$\hat{\theta} = X/n$
Exponential	λ	$\hat{\lambda} = 1/\bar{X}$
Poisson	λ	$\hat{\lambda} = \bar{X}$
Normal(μ, σ^2 known)	μ	$\hat{\mu} = \bar{X}$
Normal(μ known)	σ^2	$\hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \mu)^2$