Probability and Statistics Summary

Conditional probability

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Addition rule

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Multiplication rule

$$\Pr(A \cap B) = \Pr(B) \times \Pr(A|B)$$

Law of total probability

$$\Pr(A) = \Pr(B) \times \Pr(A|B) + \Pr(B') \times \Pr(A|B')$$

Complement

$$\Pr(A') = 1 - \Pr(A)$$

Combinations

$${}^{n}\mathrm{C}_{r} = \binom{n}{r} = \frac{n!}{r!\,(n-r)!},$$ or Pascal's Triangle

Tree diagram: use for sampling / conditional probability questions

Karnaugh map (table): use for intersection and union questions

Discrete random variable

Probability density given by a table / graph / function

$$\mu = \sum x \Pr(X = x), \qquad \sigma^2 = \sum (x-\mu)^2 \Pr(X = x) = \sum x^2 \Pr(X = x) - \mu^2$$

mean $\mu = \mathrm{E}(X)$ variance $\mathrm{var}(X) = \sigma^2 = \mathrm{E}\big((X-\mu)^2\big) = \mathrm{E}(X^2) - \mu^2$

Binomial distribution

$X \sim \mathrm{Bi}(n, p), \qquad x \in \{0, 1, \ldots, n\}$

$$\Pr(X = x) = {}^{n}\mathrm{C}_{x} p^{x} (1-p)^{n-x}$$

$$\mu = np, \qquad \sigma^2 = np(1-p)$$

Sample proportions

$\hat{P} = \frac{X}{n}, \qquad X \sim \mathrm{Bi}(n, p)$

$$\Pr(\hat{P} = \hat{p}) = {}^{n}\mathrm{C}_{x} p^{x} (1-p)^{n-x}$$

$$\mu = p, \qquad \sigma^2 = \frac{p(1-p)}{n}$$

Continuous random variable

Probability density given by function

$$\Pr(a \le X \le b) = \int_a^b f(x)\mathrm{d}x$$

$$\mu = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x \quad \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x - \mu^2 \qquad \int_{-\infty}^{m} f(x)\mathrm{d}x = \frac{1}{2}$$

Normal distribution

$X \sim \mathrm{N}(\mu, \sigma)$

CAS / 68-95-99.7% Rule

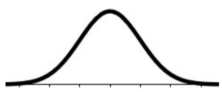
Standard normal distribution

$Z \sim \mathrm{N}(0,1), \qquad z = \frac{x-\mu}{\sigma}$

Sample proportions

$$\hat{P} \sim \mathrm{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Approximately 68% of the data lies within 1 standard deviation
Approximately 95% of the data lies within 2 standard deviations
Approximately 99.7% of the data lies within 3 standard deviations



Confidence interval

$$\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right),$$

for 90%, $z \approx 1.64$
for 95%, $z \approx 1.96$ or $z \approx 2$
for 99%, $z \approx 2.58$