MSE201 Statistics Formula Sheet

1. For the **sample space** S, the empty set ϕ, and events A and B:

$$P(S) = 1, \qquad P(\phi) = 0, \qquad P(A') = 1 - P(A).$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
$$(A \cup B)' = A' \cap B', \qquad (A \cap B)' = A' \cup B'.$$

   Conditional probability: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$ provided that $P(B) > 0$.

   Multiplication rule: $P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$.

   Independence: Events A and B are **independent** if $P(B \mid A) = P(A)$.

   Bayes' rule: $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)} = \dfrac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A')P(A')}$.

2. A **discrete** random variable X has the **probability mass function** $\{p_x\} = \{P(X = x)\}$.

   The **expectation**: $E(X) = \mu = \sum_x x p_x$.

   From random sample $x_1, \ldots, x_n$, the **sample mean** $\bar{x} = \frac{1}{n}\sum_k x_k$ estimates μ.

   The **variance** $\text{var}(X) = \sigma^2 = E((X - \mu)^2) = E(X^2) - \mu^2$, where $E(X^2) = \sum_x x^2 p_x$.

   The **sample variance**: $s^2 = \frac{1}{n-1}\left\{\sum_k x_k^2 - \frac{1}{n}\left(\sum_k x_k\right)^2\right\}$ estimates σ^2.

   The **standard deviation**: $\text{sd}(X) = \sigma = \sqrt{\text{var}(X)}$.

3. **Binomial distribution**: X is $Binomial(n, p)$.

$$p_x = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \ldots, n); \quad \mu = np,\ \sigma^2 = np(1-p).$$

   Poisson distribution: X is $Poisson(\mu)$.

$$p_x = \frac{\mu^x e^{-\mu}}{x!} \quad (x = 0, 1, 2, \ldots)\ \text{(with } \mu > 0); \quad \mu = \ \sigma^2 = \mu.$$

4. For **continuous** random variables:

   The **cumulative distribution function** (cdf)

$$F(x) = P(X \le x) = \int_{x_0=-\infty}^{x} f(x_0)\, \mathrm{d}x_0$$

   The **probability density function** (pdf) $f(x) = \dfrac{\mathrm{d}F(x)}{\mathrm{d}x}$.

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x, \quad E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, \mathrm{d}x, \quad \sigma^2 = E(X^2) - \mu^2.$$

5. **Uniform distribution** $Uniform(\alpha, \beta)$:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & (\alpha \le x \le \beta), \\ 0 & (\text{otherwise}). \end{cases} \qquad \mu = (\alpha + \beta)/2, \quad \sigma^2 = (\beta - \alpha)^2/12.$$

Exponential distribution $Exponential(\lambda)$:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \ge 0), \\ 0 & (\text{otherwise}). \end{cases} \qquad \mu = 1/\lambda \quad \sigma^2 = 1/\lambda^2.$$

Normal distribution $N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (-\infty < x < \infty); \quad E(X) = \mu, \;\; \mathrm{var}(X) = \sigma^2.$$

Standard normal distribution $N(0,1)$: If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma}$ is $N(0,1)$. For Z we write $\phi(z)$ for the pdf $f(z)$ and $\Phi(z)$ for the cdf $F(z)$.

6. For a system of k devices, which operate independently:

   The **system reliability**, R is the probability of a path of operating devices.

   Let $R_i = P(D_i) = P(\text{"device } i \text{ operates"})$.

   A system of devices in **series** fails if any device fails.

   $R = P(D_1 \cap D_2 \cap \ldots \cap D_k) = R_1 R_2 \ldots R_k$.

   A system of devices in **parallel** operates if any device operates.

   $R = P(D_1 \cup D_2 \cup \ldots \cup D_k) = 1 - (1 - R_1)(1 - R_2) \ldots (1 - R_k)$.

7. If $x_1, \ldots, x_n$ are a random sample from $N(\mu, \sigma^2)$ and σ^2 is known, then

   the $100(1-\alpha)\%$ CI for μ is $(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$,

   where $z_{\alpha/2}$ is that value such that $1 - \alpha = P(-z_{\alpha/2} < Z < z_{\alpha/2})$, where $Z \sim N(0,1)$.

   If σ^2 is estimated, then the $100(1-\alpha)\%$ CI for μ is $(\bar{x} - t^{n-1}_{\alpha/2} s/\sqrt{n}, \bar{x} + t^{n-1}_{\alpha/2} s/\sqrt{n})$

   where $t^{n-1}_{\alpha/2}$ is that value such that $1 - \alpha = P(-t^{n-1}_{\alpha/2} < T < t^{n-1}_{\alpha/2})$,

   where T has a t-distribution with $n - 1$ degrees of freedom.

   A **significance test** of H_0 rejects H_0, if, assuming that H_0 is true, a test statistic is in a rejection region of its sampling distribution.