BST 140.651 Formula sheet

1. **DeMorgan's laws** state that $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

2. **Baye's rule**
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

3. **Binomial** mass function $P(X = k) = \left(\begin{array}{c} n \\ k \end{array}\right) p^k(1-p)^{n-k}$ for $x = 0, \ldots, n$

4. **The normal pdf** $f(x) = (2\pi\sigma^2)^{-1/2} \exp\{-(x-\mu)^2/2\sigma^2\}$ for $-\infty < x < \infty$.

5. **Std. Z quantiles** $Z_{.90}$, $Z_{.95}$, $Z_{.975}$ and $Z_{.99}$ are 1.28, 1.645, 1.96 and 2.32, respectively.

6. **Sample variance** of data $x_1, \ldots, x_n$ with sample mean $\bar x$ is
$$\frac{\sum_{i=1}^n (x_i - \bar x)^2}{n-1}$$

7. $\sum_{i=1}^n (x_i - \bar x)^2 = \sum_{i=1}^n x_i^2 - n\bar x^2$

8. **variance** of a random variable X is $\sigma^2 = E\left[(X-\mu)^2\right] = E[X^2] - E[X]^2$.

9. **Chebyshev's inequality** states that for any random variable $P(|X - \mu| \geq K\sigma) \leq 1/K^2$.

10. **Central Limit Theorem**. If the X_i are iid with mean μ and (finite) variance σ^2 then
$$Z = \frac{\bar X - \mu}{\sigma/\sqrt{n}}$$
will limit to a standard normal distribution.

11. Assuming normality and a sample standard deviation of S $\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]$ is a $100(1-\alpha)\%$ confidence interval for σ^2

12. If Z is standard normal and X is and independent Chi-squared with df degrees of freedom then $\frac{Z}{\sqrt{X/df}}$ follows a Student's T distribution with df degrees of freedom.

13. The **Score interval** is obtained by inverting a score test
$$\hat p \left(\frac{n}{n+Z^2_{1-\alpha/2}}\right) + \frac{1}{2}\left(\frac{Z^2_{1-\alpha/2}}{n+Z^2_{1-\alpha/2}}\right)$$
$$\pm Z_{1-\alpha/2}\sqrt{\frac{1}{n+Z^2_{1-\alpha/2}}\left[\hat p(1-\hat p)\left(\frac{n}{n+Z^2_{1-\alpha/2}}\right) + \frac{1}{4}\left(\frac{Z^2_{1-\alpha/2}}{n+Z^2_{1-\alpha/2}}\right)\right]}.$$

14. For independent groups of iid variables, say X_i and Y_i, *with a constant variance* σ^2 across groups

$$Z = \frac{\bar X - \bar Y - (\mu_x - \mu_y)}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

limits to a standard normal random variable as both n_x and n_y get large. Here

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

is the **pooled estimate** of the variance. Obviously, $\bar X$, S_x, n_x are the sample mean, sample standard deviation and sample size for the X_i and $\bar Y$, S_y and n_y are defined analogously.

15. The statistic

$$\frac{\bar Y - \bar X - (\mu_y - \mu_x)}{\left(\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)^{1/2}}$$

approximately follows Gosset's T distribution with degrees of freedom equal to

$$\frac{\left(S_x^2/n_x + S_y^2/n_y\right)^2}{\left(\frac{S_x^2}{n_x}\right)^2/(n_x - 1) + \left(\frac{S_y^2}{n_y}\right)^2/(n_y - 1)}$$