Probability Formula Review

I. Types and characteristics of probability

A. Types of probability
1. Classical: \( P(A) = \frac{A}{N} \)
2. Empirical: \( P(A) = \frac{A}{n} \)
3. Subjective: Use empirical formula assuming past data of similar events is appropriate.

B. Probability characteristics
1. Range for probability: \( 0 \leq P(A) \leq 1 \)
2. Value of complements: \( P(\bar{A}) = 1 - P(A) \)

II. Probability rules

A. Addition is used to find the sum or union of 2 events.
1. General rule: \( P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \)
2. Special rule: \( P(A \text{ or } B) = P(A) + P(B) \) is used when events are mutually exclusive.

B. Multiplication is used to determine joint probability or the intersection of 2 events.
1. General rule: \( P(A \text{ and } B) = P(A) \times P(B \mid A) \)
2. Special rule: \( P(A \text{ and } B) = P(A) \times P(B) \) is used when the events are independent.

Note: For independent events, the joint probability is the product of the marginal probabilities.

C. Bayes' theorem is used to find conditional probability.
\[
P(A \mid B) = \frac{P(A) \times P(B \mid A)}{P(A) \times P(B \mid A) + P(\bar{A}) \times P(B \mid \bar{A})}
\]

Note: The denominator is when condition \( B \) happens. It happens with \( A \) and with \( \bar{A} \).

III. Counting rules

A. The counting rule of multiple events: If one event can happen \( M \) ways and a second event can happen \( N \) ways, then the two events can happen \( M \times N \) ways. For 3 events, use \( M \times N \times O \).

B. Factorial rule for arranging all of the items of one event: \( N \) items can be arranged in \( N! \) ways.

C. Permutation rule for arranging some of the items of one event:
(order is important: a, b, c and c, a, b are different)
\[
N^P_R = \frac{N!}{(N-R)!}
\]

D. Combination rule for choosing some of the items of one event:
(order is not important: abc and cba are the same and are not counted twice)
\[
N^C_R = \frac{N!}{(N-R)! \times R!}
\]

IV. Discrete probability distributions

A. Probability distributions
1. \( P(x) = [x \times P(x)] \) is calculated for each value of \( x \).
2. Mean of a probability distribution: \( \mu = E(x) = \Sigma [x \times P(x)] \)
3. Variance of a probability distribution: \( V(x) = [\Sigma x^2 \times P(x)] - [E(x)]^2 \)

B. Binomial distributions
\[
P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}
\]
where
\( n \) is number of trials \( x \) is number of successes
\( p \) is probability of success \( q \), the probability of failure, is \( 1 - p \)
\( \mu = np, \sigma^2 = npq \) and \( \sigma = \sqrt{npq} \)

C. Poisson distributions
\[
P(x) = \frac{e^{-\mu} \mu^x}{x!}
\]
where \( \mu = np \) Poisson approximation of the binomial requires \( n \geq 30 \) and \( np < 5 \) or \( nq < 5 \).
V. The continuous normal probability distribution

A. To find the probability of $x$ being within a given range: 
   $$Z = \frac{x-\mu}{\sigma}$$ 
   Normal approximation of the binomial requires $n \geq 30$ and both $np$ and $nq$ are $\geq 5$. The continuity correction factor applies.

B. To find a range for $x$ given the probability: $\mu \pm Z\sigma$

VI. Central limit theorem

[Diagram showing sampling distribution of the means]

If $n \geq 30$, the population may be skewed.

VII. Point estimates

A. $\bar{x}$ for $\mu$  
B. $s$ for $\sigma$  
C. $\bar{p}$ for $p$  
D. $s_{\bar{x}}$ for $\sigma_{\bar{x}}$ where $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

VIII. Interval estimates when $n \geq 30$

A. For a population mean $\bar{x} \pm Z\frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm Z\frac{s}{\sqrt{n}}$

B. For a population proportion $\bar{p} \pm Z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ where $\bar{p} = \frac{x}{n}$

IX. Determining sample size

A. When estimating the population mean $n = \left(\frac{Z_\alpha E}{E}\right)^2$

B. When estimating the population proportion $n = \frac{\bar{p}(1-\bar{p})}{\left(\frac{Z_\alpha}{E}\right)^2}$

[Note: Use the finite correction factor in section VIII formulas when $n/N \geq .05$. $n = \frac{N-n}{N-1}$]

Section VIII Note: When $n < 30$ and $\sigma$ is unknown, the $t$ distribution, to be discussed in chapter 16, must be substituted for the $z$ distribution when making interval estimates. Many statistics software programs do all interval calculations, regardless of sample size, using the $t$ distribution.