MATH 147B Exam 2 Study Guide

David Collins

October 27, 2022

Note 1: This guide is not all-encompassing. It's just something I put together to put all of the million formulas you need to know in one place.

Note 2: $\cap$ =and, $\cup = or$

1 Probability

There are five rules of probability:

1) For any event A, $0 \leq P(A) \leq 1$.

2) Let S be the sample space of events. Then $P(S) = 1$.

3) Complement rule: $P(A^c) = 1 - P(A)$

4) Addition rule: $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint.

5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ if A and B are independent.

General Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ EVEN IF A and B are independent.

Condition probability: $P(B|A) = \frac{P(A \cap B)}{P(A)}$.

General Multiplication Rule: $P(A \cap B) = P(A)P(B|A)$.

Events A and B are independent if $P(B|A) = P(B)$.

Know how to make and use tree diagrams!

2 Random Variables

A random variable's numeric value is based on the outcome of a random event.

A random variable is discrete if you can count all the outcomes (binomial, geometric, uniform).

A random variable is continuous if it takes on infinitely many values (Normal, exponential).

2.1 Expected Value

Expected Value is the mean, μ, of a random variable.

We say $E(X) = \mu = \sum xP(x)$.

Furthermore for real numbers a, b,

$$E(aX \pm Y \pm b) = aE(X) \pm E(Y) \pm b$$

2.2 Variance

Variance of a random variable is the expected value of the square deviation from the mean and conceptually is the same as variance from earlier chapters.

$$\sigma^2 = Var(X) = SD(X)^2 = \sum (x-\mu)^2 P(x) = E(X-\mu)^2$$

If X and Y are independent,

$$Var(X \pm Y) = Var(X) + Var(Y)$$

Also,

$$Var(aX \pm b) = a^2 Var(X)$$

3 Probability Models

A Bernoulli trial has two outcomes with the same probability, p, for each trial. Each trial is independent. Common examples are flipping a coin or rolling a die.

3.1 Geometric

The Geometric model is specified by parameter p, the probability of success, and tells you how long it will take to achieve the first success in a series of Bernoulli trials.
For Geom(p):

$$P(X = x) = pq^{x-1}, \quad \text{where q=1-p.}$$

$$E(X) = \frac{1}{p}, Var(X) = \frac{q}{p^2}$$

If the individual trials are not independent, we can proceed if the sample is smaller than 10% of the population.

3.2 Binomial

A Binomial model describes the number of successes in a specified number of trials. It takes two parameters: number of trials n and probability of success p.

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad \text{where } \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$E(X) = np,\ Var(X) = npq$$

If $np \geq 10$ AND $nq \geq 10$ then we can assume that the Binomial model is approximately Normal.

3.3 Continuity Correction

If we try to use the Normal model to approximate a Binomial probability, for say $P(X = 5)$, we do $P(4.5 \leq X \leq 5.5)$ on the Normal model to get a close value.

3.4 Exponential Model

This is for continuous random variables, like time. $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0, \lambda > 0$.

4 Sampling Distributions and Confidence Intervals for Proportions

The distribution of the statistics over all possible samples is the sampling distribution.
The standard error for a proportion $\hat{p}$ is

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The critical value z^* is the z-score that corresponds to a percentage for the confidence interval. Thus, the confidence interval is

$$(\hat{p} - z^* SE(\hat{p}), \hat{p} + z^* SE(\hat{p}))$$

4.1 IMPORTANT

To interpret a confidence interval, say "We are X% confident that between (lower value)% and (upper value)% of (Random Variables) are (something)." In general, you never know if the true mean/proportion is in your confidence interval. HOWEVER, if you took lots and lots of confidence intervals, you would expect the true mean to be contained in X% of them.

4.2 Finding z^* and t^*

Consider $\alpha =$ the percentage of how confident we want to be. For example, 95% confidence corresponds to $\alpha = .95$. To find z^*, we find the area under the normal curve equal to $1 - \frac{1-\alpha}{2}$. For example, $1 - \frac{1-.95}{2} = 1 - .25 = .975$. Then, find the area of .9750 (or closest to it) and then backsolve for the z-score. For the example of 95% confidence, the z-score should be 1.96, which we round to 2.
For a $t-$score, we will be given the confidence level and degrees of freedom, since $df = n - 1$, the sample size minus one. Go up and across to find the given value.

5 Confidence Interval for Means

Central Limit Theorem: When a random sample is drawn from a population with mean μ and standard deviation σ, the sample mean $\bar{y}$ has sampling distribution $N(\mu, \frac{\sigma}{\sqrt{n}})$ for a sufficiently large sample size n.

5.1 Student's t distribution

This distribution is very dependent on sample size, and it changes with the sample size, but is always bell shaped. As the sample size gets very large, it becomes nearly Normal. Degrees of freedom is $df = n - 1$, the sample size minus one.
The t-score (similar to z-score) is

$$t = \frac{\bar{y} - \mu}{SE(\bar{y})}$$

Thus the confidence interval for a t model is

$$(\bar{y} - t^*_{n-1} SE(\bar{y}), \bar{y} + t^*_{n-1} SE(\bar{y})), \quad \text{where } SE(\bar{y}) = \frac{s}{\sqrt{n}}$$

To use this model, the data values should be independent, from a random sample, and from a unimodal and symmetric data set.

5.2 Bootstrapping

We can resample from the same sample group to create more trials.