Calculus, Probability, and Related Material for Stat 343

I'm going to try to collect together some things from Calculus and related topics that will be useful for Stat 343. Let me know if you see any mistakes come across anything else that it would be helpful to add to this.

Miscellaneous

Logarithms and Exponents

a, b, and c are real numbers, $e \approx 2.718281828459$ is Euler's number.

$\log(a)$ is defined to be the number to which you raise e in order to get a: $e^{\log(a)} = a$.

$\log(e) = 1$

$\log(ab) = \log(a) + \log(b)$

$\log(a^b) = b \log(a)$

$\log(a/b) = \log(a) - \log(b)$

$a^b \cdot a^c = a^{b+c}$

Gamma Function

The Gamma (Γ) function is basically a continuous version of the factorial. If a is an integer, then

$\Gamma(a) = (a - 1)!$

If a is a real number, the Γ function is still defined, and it's basically a smooth interpolation between the factorials of nearby integers. There's a way to define the Γ function as an integral, but we won't need to know that.

Probability

Let X, Y be random variables whose distributions depends on parameters θ_X and θ_Y respectively, and use lower case x and y to denote observed values or values we might hypothetically observe.

Probability Mass/Density Function (p.m.f., p.d.f.), Cumulative Distribution Function (c.d.f.):

If X is discrete, then the probability mass function $f(x) = P(X = x)$

If X is continuous, then the probability density function $f(x)$ can be used to find $P(X \in [a, b]) = \int_a^b f(x)dx$

The cumulative distribution function is $F(x) = P(X \le x)$:

- If X is discrete, $F(t) = \sum_{t=-\infty}^x f(t)$
- If X is continuous, $F(t) = \int_{-\infty}^x f(t)dt$

Joint Distributions from the Marginals

If X and Y and both discrete, then they have a joint p.m.f.: $f(x, y) = P(X = x \text{ and } Y = y)$

If X and Y are both continuous, then they have a joint p.d.f.: $P(X \in [a, b] \text{ and } Y \in [c, d]) = \int_a^b \int_c^d f(x, y) \, dx \, dy$

If one of X and Y is discrete and the other is continuous, it's possible to define a similar probability function $f(x, y)$.

If X and Y are independent, then their joint p.f. is the product of their marginal p.f.'s:

$$f(x,y) = f(x)f(y)$$

If X and Y are **not** independent, their joint p.f. is the product of the marginal for one and the conditional for the second given the first:

$$f(x,y) = f(x)f(y|x) = f(y)f(x|y)$$

Marginal distributions from the Joint

Suppose X and Y have joint probability function $f(x,y)$.

If X is discrete, then the marginal probability function for Y is

$$f(y) = \sum_x f(x,y)$$

If X is continuous, then the marginal probability function for Y is

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Conditional Distributions

By definition, the conditional distribution for $Y|X = x$ has p.f.

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

Bayes' Rule

If I know the marginal distribution for X has p.f. $f(x)$ and the conditional distribution for $Y|X$ has p.f. $f(y|x)$ then I can calculate the p.f. for the conditional distribution of $X|Y$ as follows:

$$
\begin{aligned}
f(x|y) &= \frac{f(x,y)}{f(y)} \\
&= \frac{f(x,y)}{\int f(x,y)dx} \\
&= \frac{f(x)f(y|x)}{\int f(x)f(y|x)dx}
\end{aligned}
$$

If X is discrete, replace the integral in the denominator by a summation.

There are two ways of explaining why this is useful:

1. It lets us reverse the order of conditioning from $Y|X$ (what we know to start with) to $X|Y$.

2. It lets us update our knowledge about the distribution of X having observed a value of $Y = y$.

Expected Value and Variance

$E(X) = \int x f(x) dx$

$$
\begin{aligned}
Var(X) &= \int (x - E(X))^2 f(x) dx \\
&= \int (x^2 - 2xE(X) + E(X)^2) f(x) dx \\
&= \int x^2 f(x) dx - 2E(X) \int x dx + E(X)^2 \int f(x) dx \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

$E(aX + b) = aE(X) + b$

$Var(aX + b) = a^2 Var(X)$

Differential Calculus

Chain rule:

Suppose f and g are functions, and define F by $F(x) = f(g(x))$. Then $F'(x) = f'(g(x)) \cdot g'(x)$

Derivative of a polynomial:

If $a \neq 0$ then $\frac{d}{dx} x^a = ax^{a-1}$

Two special cases that come up a lot are $a = -1$ and $a = -2$:

If $a = -1$ then $\frac{d}{dx} \frac{1}{x} = \frac{d}{dx} x^{-1} = -1x^{-2} = \frac{-1}{x^2}$

If $a = -2$ then $\frac{d}{dx} \frac{1}{x^2} = \frac{d}{dx} x^{-2} = -2x^{-3} = \frac{-2}{x^3}$

Derivative of an exponential:

$\frac{d}{dx} e^x = e^x$

In combination with the chain rule, we get

$\frac{d}{dx} e^{f(x)} = e^{f(x)} f'(x)$

Derivative of a logarithm:

$\frac{d}{dx} \log(x) = \frac{1}{x}$

In combination with the chain rule, we get

$\frac{d}{dx} \log(f(x)) = \frac{1}{f(x)} f'(x)$

Taylor's Theorem

I adapted this statement of Taylor's theorem from Wikipedia: https://en.wikipedia.org/wiki/Taylor%27s_theorem#Taylor's_theorem_in_one_real_variable

Let $k \geq 1$ be an integer and suppose that the function $f : \mathbb{R} \to \mathbb{R}$ is k times differentiable at the point $a \in \mathbb{R}$. Define the k-th order polynomial approximation to f centered at a by

$P_k(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x - a)^k$

Then there exists a function $h_k : \mathbb{R} \to \mathbb{R}$ such that:

- $f(x) = P_k(x) + h_k(x)(x-a)^k$ and
- $\lim_{x \to a} h_k(x) = 0$

(You can get more specific about what the function h_k looks like and rates of convergence to 0, but we don't need to do that.)

The main point is that for values of x near a, the function $f(x)$ can be well approximated by a polynomial, and the polynomial's coefficients can be obtained by the derivatives of f.

As an example, let's approximate $f(x) = e^{5x}$ by a second-order Taylor polynomial centered at $a = 1$. We will need the first and second derivatives of $f(x)$:

$$\frac{d}{dx}e^{5x} = e^{5x} \cdot 5$$

$$\frac{d^2}{dx^2}e^{5x} = \frac{d}{dx}e^{5x} \cdot 5 = e^{5x} \cdot 25$$

$$P_2(x) = f(1) + f'(1)(x-1) + \frac{f''(1)}{2}(x-1)^2$$

$$= e^{5 \cdot 1} + 5e^{5 \cdot 1}(x-1) + \frac{25e^{5 \cdot 1}}{2}(x-1)^2$$

The claim is that $f(x)$ looks very similar to $P_2(x)$ for values of x near $a = 1$. Let's verify with a picture:

```
library(ggplot2)
f <- function(x) {
  exp(5 * x)
}

P_2 <- function(x) {
  exp(5) + 5 * exp(5) * (x - 1) + (25 * exp(5)) / 2 * (x - 1)^2
}

temp_df <- data.frame(x = c(0.5, 1.5))
ggplot(data = temp_df, mapping = aes(x =x)) +
  stat_function(fun = f) +
  stat_function(fun = P_2, color = "orange")
```