10-601B Recitation 1

Calvin McCarter

September 3, 2015

1 Probability

1.1 Linearity of expectation

For any random variable X and constants a and b:

$$\mathbb{E}[a + bX] = a + b\,\mathbb{E}[X]$$

For any random variables of X and Y, whether independent or not:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Recall the definition of variance:

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Now let's define $Y = a + bX$ and show that $\mathrm{Var}[Y] = b^2\,\mathrm{Var}[X]$:

$$\mathbb{E}[Y] = a + b\,\mathbb{E}[X] \qquad \text{by linearity of expectation}$$

Now we can derive the variance:

$$\begin{aligned}
\mathrm{Var}[Y] &= \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] && \text{definition of variance} \\
&= \mathbb{E}\left[\left([a + bX] - [a + b\,\mathbb{E}\,X]\right)^2\right] \\
&= \mathbb{E}\left[b^2(X - \mathbb{E}\,X)^2\right] \\
&= b^2\,\mathbb{E}\left[(X - \mathbb{E}\,X)^2\right] && \text{linearity of expectation} \\
&= b^2\,\mathrm{Var}[X] && \text{definition of variance}
\end{aligned}$$

This is why we often use the standard deviation (the square root of variance), because $\mathrm{StdDev}[Y] = b\,\mathrm{StdDev}[X]$, which is more intuitive.

1.2 Prediction, and expectation, and partial derivatives

Suppose we want to predict a random variable Y simply using some constant c. What value of c should we choose? Here we show that $\mathbb{E}[Y]$ is a sensible choice.

But first, we need to decide what a good prediction should look like. A common choice is the mean-squared error, or MSE. We punish our prediction ever more harshly the further it gets from the observed Y.

$$\text{MSE} = \mathbb{E}\left[(Y - c)^2\right]$$

We now show that MSE is minimized at $\mathbb{E}[Y]$. We set it up as an optimization problem:

$$\begin{aligned} &\min_c \mathbb{E}\left[(Y-c)^2\right] \\ =&\min_c \mathbb{E}\left[Y^2 - 2\,\mathbb{E}[Y]c + c^2\right] \\ =&\min_c \mathbb{E}[Y^2] - 2\,\mathbb{E}[Y]c + c^2 \end{aligned}$$

This is a quadratic function of c. We can find the minimum of this quadratic by setting its partial derivative to 0, and solving for c:

$$\begin{aligned} \frac{\partial}{\partial c}\left[\mathbb{E}[Y^2] - 2\,\mathbb{E}[Y]c + c^2\right] =& 0 \\ -2\,\mathbb{E}[Y] + 2c =& 0 \\ c =& \mathbb{E}[Y] && \text{This minimizes the MSE!} \end{aligned}$$

1.3 Sample mean and the Central Limit Theorem

Suppose we have n random variables $X_1, ..., X_n$ that are independent and identically distributed (iid). Suppose we don't know what the distribution is, but we do know their expectation and variance:

$$\mathbb{E}[X_i] = \mu \text{ and } \operatorname{Var}[X_i] = \sigma^2 \quad \text{for } i = 1, ..., n$$

A common way to estimate the unknown μ is to use the average (sample mean) of our data:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

How does this estimate behave? We can characterize its behavior by deriving its expectation and variance.

$$\begin{aligned} \mathbb{E}[\bar{X}_n] =& \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] \\ =& \frac{\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]}{n} && \text{linearity of expectation} \\ =& \frac{n\mu}{n} = \mu \end{aligned}$$

This tells us that $\bar{X}_n$ is "unbiased" - its expected value is the true mean.

$$\begin{aligned}
\mathrm{Var}[\bar{X}_n] =& \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] \\
=& \frac{1}{n^2}\,\mathrm{Var}\left[X_1 + \cdots + X_n\right] \\
=& \frac{1}{n^2}\Big(\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]\Big) \quad \text{only because } X_i \text{ are iid - variance isn't linear!} \\
=& \frac{1}{n^2}(n\,\mathrm{Var}[X_i]) = \frac{\sigma^2}{n}
\end{aligned}$$

This tells us that the variance of the average decreases as n the number of samples increases.

But it turns out we know something more about the distribution of $\bar{X}_n$. It's distribution actually converges to a Normal distribution as n gets large. This is called the Central Limit Theorem:

$$\bar{X}_n \rightsquigarrow \mathcal{N}\Big(\mu, \frac{\sigma^2}{n}\Big)$$

2 Linear Algebra

I discussed problems taken directly from Section 4 of Linear Algebra Review. Two other great online resources:

- YouTube tutorial on gradients
- Matrix Cookbook reference