STAT 3503/8109 Lecture 4 Notes

Edoardo Airoldi

*Scribe: Srikar Katta**

Fall 2020: September 21, 2020

1 Introduction

Probability theory is essential to the study of statistics and data science. Last week, we saw that idea with marginalizing joint probability distributions using integrals. This week, we will see the properties of the expectation and variance operators. Namely, we will see the following:

- Properties of $\mathbb{E}$
- Properties of $\mathbb{V}$
- $\mathbb{E}[a + bX] = a + b\mathbb{E}[X]$
- $\mathbb{V}[a + bx] = b^2\mathbb{V}[x]$
- Using the *delta method* to approximate non-linear functions of x.

2 Properties of Expectation Operator

Suppose $X \sim f_X(x) \equiv X \sim f_X(x|\theta) \equiv X \sim f_{X|\theta}(x)$. We are interested in finding the following:

*Please share any comments or suggestions with Srikar Katta at srikar@temple.edu

- $\mathbb{E}[X]$
- $\mathbb{E}[a = bX]$
- $\mathbb{E}[g(x)]$, where g is some non-linear function

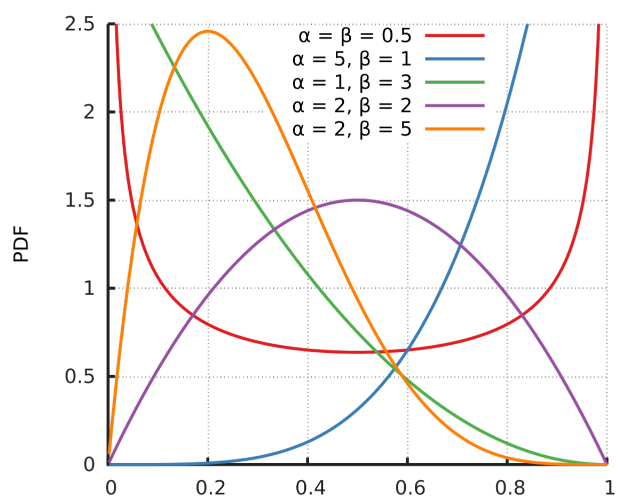
We will see the third in its own section in conjunction with the same idea for the variance.

2.1 $\mathbb{E}[X]$

Suppose $x \sim Normal(x|\mu, \sigma^2)$. We know X is parameterized by the mean and variance so $\mathbb{E}[X] = \mu$. Given $f_X(x|\theta)$ is a standard distribution[1], then $\mathbb{E}[X] = function(\theta)$, where θ represents the parameters of our standard distribution.

Example 2.1. *Let some random variable*[2] $X \in [0, 1]$ *follow a Beta distribution:* $X \sim Beta(x|\alpha, \beta)$. *What is* $\mathbb{E}[X]$*?*

First, we know our density functions could look like one of the following because X follows a *Beta* distribution:



So,

$$\mathbb{E}[X] = \int_{Support of X} x * f_X(\theta)dx$$

[1]What is considered a standard distribution can be found in Appendix

[2]Recall that every random variable is defined by two things: the support (the set of possible values it could adopt) and its distribution

$$
\begin{aligned}
&= \int_0^1 x * f_X(\theta) dx \\
&= \int_0^1 \frac{\Gamma(\alpha)}{\Gamma(\alpha+\beta)} * x^{\alpha-1} * (1-x)^{\beta-1} * x dx, \ (^3) \\
&= \frac{\alpha}{\alpha+\beta}.
\end{aligned}
$$

Notice, Γ is just some function and we replace $f_X(\theta)$ with the functional form of our probability distribution. And from integrating our distribution (the exact steps for which is not important), we see that $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$.

2.2 $\mathbb{E}[a + bX]$

Now, we can show the linearity of the expectation operator. We can prove this general claim via our normal proof techniques.

Proof: Assume some random variable X whose support is some subset of the real numbers, (a, b) has the following generic probability distribution: $X \sim f_x(x|\theta)$ with support S. Then,

$$
\begin{aligned}
\mathbb{E}[a+bX] &= \int_{Support\ of\ X} (a+bx) * f_x(x|\theta) dx \\
&= \int_S (a+bx) * f_x(x|\theta) dx \\
&= \int_S a * f_x(x|\theta) dx + \int_S bx * f_x(x|\theta) dx \\
&= a \int_S f_x(x|\theta) dx + b \int_S x * f_x(x|\theta) dx \\
&= a * 1 + b \int_S x * f_x(x|\theta) dx \text{ because the integral of PDF} = 1 \\
&= a + b\mathbb{E}[X] \text{ because } \int_S x * f_x(x|\theta) = \mathbb{E}[X].
\end{aligned}
$$

Therefore, the expectation operator is linear. $\square$

[3] $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ or the special case for $z \in \mathbb{Z}, \Gamma(z) = (z-1)!$

3 Properties of $\mathbb{V}(X)$

Again, suppose $X \sim f_X(x) \equiv X \sim f_X(x|\theta) \equiv X \sim f_{X|\theta}(x)$. We are interested in finding the following:

- $\mathbb{V}[X]$
- $\mathbb{V}[a = bX]$
- $\mathbb{V}[g(x)]$, where g is some non-linear function

We will see the third in its own section in conjunction with the same idea for the variance.

3.1 $\mathbb{V}[X]$

As earlier, assume $X \sim f_X(x|\theta)$ and $X \in (a,b)$ where (a,b) is some subset of the real numbers. Usually, if f_X is one of those standard distributions, then we know what $\mathbb{E}[X]$ is. If not, we can compute it. But, from that, we know that the variance is the following:

$$
\begin{aligned}
\mathbb{V}[X] &= \int_{Support\ of\ X} (X - \mathbb{E}[X])^2 f_X(\theta) dx \\
&= \int_a^b (X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]) f_X(\theta) dx \\
&= \int_a^b (X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]) f_X(\theta) dx \\
&= \int_a^b X^2 f_X(\theta) dx + \int_a^b \mathbb{E}[X]^2 f_X(\theta) dx - \\
&\int_a^b 2X\mathbb{E}[X] f_X(\theta) dx \\
&= \int_a^b X^2 f_X(\theta) dx + \mathbb{E}[X]^2 \int_a^b f_X(\theta) dx - 2\mathbb{E}[X] \int_a^b X f_X(\theta) dx \text{ since } \mathbb{E}[X] \text{ is fixed(constant)} \\
&= \mathbb{E}[X^2] + \mathbb{E}[X]^2 * 1 - 2\mathbb{E}[X]\mathbb{E}[X] \\
&= \mathbb{E}[X^2] + \mathbb{E}[X]^2 * 1 - 2\mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2.
\end{aligned}
$$

So, another way of writing the variance of some random variable X is $\mathbb{E}[X^2] - \mathbb{E}[X]^2$.$\square$

3.2 $\mathbb{V}[a + bX]$

Try this out on your own before seeing this proof written out.

Proof: Let X be a random variable with probability distribution function f_X (for ease of explanation, we do not differentiate between PDF and PMF) and support $S \subseteq \mathbb{R}$ (a subset of the real numbers). And let $a, b \in \mathbb{R}$.

$$
\begin{aligned}
\mathbb{V}[a + bX] &= \int_S ((a + bx) - \mathbb{E}[a + bX])^2 f_X dx \\
&= \int_S ((a + bx) - (a + b\mathbb{E}[X]))^2 f_X dx \text{ – properties of expectation} \\
&= \int_S \big((a + bx)^2 + (a + b\mathbb{E}[X])^2 - 2(a + bx)(a + b\mathbb{E}[X])\big) f_X dx \text{ – FOIL method} \\
&= \int_S \big(a^2 + b^2x^2 + 2abx + a^2 + b^2\mathbb{E}[X]^2 + 2ab\mathbb{E}[X] - 2(a^2 + ab\mathbb{E}[X] + abx + b^2\mathbb{E}[X])\big) f_X dx \\
&= \int_S \big(2a^2 + b^2x^2 + 2abx + b^2\mathbb{E}[X]^2 + 2ab\mathbb{E}[X] - 2a^2 - 2ab\mathbb{E}[X] - 2abx - 2b^2\mathbb{E}[X])\big) f_X dx \\
&= \int_S \big(2a^2 - 2a^2 + b^2x^2 + 2abx - 2abx + b^2\mathbb{E}[X]^2 - 2b^2\mathbb{E}[X] + 2ab\mathbb{E}[X] - 2ab\mathbb{E}[X])\big) f_X dx \\
&= \int_S \big(b^2x^2 - b^2\mathbb{E}[X]^2 +)\big) f_X dx \\
&= b^2 \int_S \left(x^2 - \mathbb{E}[X]^2\right) f_X dx \\
&= b^2 \int_S x^2 f_X dx - \int_S \mathbb{E}[X]^2 f_X dx \\
&= b^2 \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2 \int_S f_X dx\right) \\
&= b^2 \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) \\
&= b^2 \mathbb{V}[X],
\end{aligned}
$$

our desired result. Thus, $\mathbb{V}[a + bX] = b^2\mathbb{V}[X]$. □

4 Delta Method

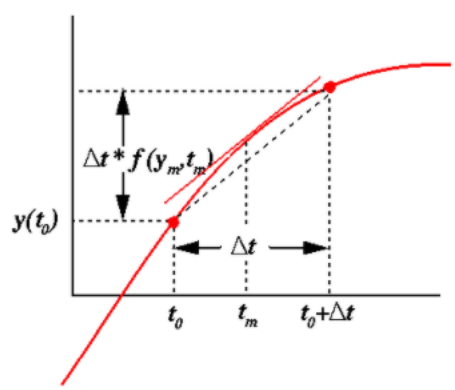
From our calculus courses, we know that g can be approximated by Taylor series expansion. We can rewrite the first order Taylor expansion by using the definition of a first derivative:

$$\begin{aligned} g'(x) &= \lim_{h\to 0} \frac{g(x_0+h)-g(x_0)}{h} \\ \equiv g'(x) &= \lim_{x\to x_0} \frac{g(x)-g(x_0)}{x-x_0}, \text{ which implies} \\ g'(x) &\approx \frac{g(x)-g(x_0)}{x-x_0} \end{aligned}$$

So, with some algebraic manipulation, we see that

$$g(x) \approx g(x_0) + g'(x_0)(x - x_0),$$

which can be visualized as follows:



In words, this is saying that we can approximate a non-linear function using the tangent line (aka the derivative at a certain point) and just go down the slope by the number of units we want (this is what we do in $g'(x_0)(s - s_0)$. Now, we can apply this to problems in statistics as well.

Let us assume that some random variable X has some density function $f_X(x|\theta)$. Assume, for convenience $\theta = \mathbb{E}[X]$. Let the random variable $Y = g(X)$, where g is some nonlinear function (e.g., $y = e^x, y = \frac{1}{x}, y = ln(x)$). What is $\mathbb{V}[Y]$? For this, we need to find the density of Y. Let's say that $X \sim Normal(\mu, \sigma^2)$ and $y = x^2$, so the function g that sends x to y is a non-linear function. Then, since the χ^2 distribution is the "square" of a normal distribution, Y would follow a χ^2 distribution. Let's illustrate the steps to finding this with an example.

4.1 First Order Taylor Expansion to Find Expectation

Example 4.1. *Let $X \sim f_X(x|\theta)$, where $x \in (0, \infty)$. Let $y = g(x) = \frac{1}{x}$. What is $\mathbb{E}[Y] = \mathbb{E}[g(x)]$?*

So, returning to our problem, $\mathbb{E}[g(x)]$ can be rewritten as

$$\mathbb{E}[g(x)] \approx \mathbb{E}[g(x_0) + g'(x_0)(x - x_0)].$$

However, we run into a new problem: what should x_0 be? We can use $x_0 = \mathbb{E}[X]$. Then,

$$\begin{aligned}
\mathbb{E}[g(x)] &\approx \mathbb{E}[g(x_0) + g'(x_0)(x - x_0)] \\
&= \mathbb{E}[g(x_0)] + \mathbb{E}[g'(x_0)(x - x_0)] \\
&= \mathbb{E}[g(\mathbb{E}[X])] + \mathbb{E}[g'(\mathbb{E}[X])(x - \mathbb{E}[X])] \\
&= g(\mathbb{E}[X]) + g'(\mathbb{E}[X])(\mathbb{E}[x] - \mathbb{E}[X]) \text{ since } \mathbb{E}[X] \text{ is a constant} \\
&= g(\mathbb{E}[X]) + g'(\mathbb{E}[X]) * 0 \\
&= g(\mathbb{E}[X]).
\end{aligned}$$

So, in other words, the delta method is saying that $\mathbb{E}[g(x)] \approx g(\mathbb{E}[X])$.

Example 4.2. *Assume that $X \sim Normal(\mu, \sigma^2)$ with support $\mathbb{R}$ and $y = log(x)$. What is $\mathbb{E}[Y]$?*

This is very straightforward. From the delta method, $\mathbb{E}[Y] = \mathbb{E}[log(x)] = log(\mathbb{E}[X])$. □

4.2 Second Order Taylor Expansion to Find Expectation

Now, we can extend this forward to a second order Taylor expansion. Recall, that is

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2}g''(x_0)(x - x_0)^2.$$

So,

$$\mathbb{E}[g(x)] \approx \mathbb{E}[g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2}g''(x_0)(x - x_0)^2]$$

$$= \mathbb{E}[g(x_0)] + \mathbb{E}[g'(x_0)(x - x_0)] + \mathbb{E}[\frac{1}{2}g''(x_0)(x - x_0)^2]$$
$$= \mathbb{E}[g(x_0)] + \mathbb{E}[g'(x_0)(x - x_0)] + \frac{1}{2}\mathbb{E}[g''(x_0)(x - x_0)^2].$$

So, setting $x_0 = \mathbb{E}[X]$, we see

$$\begin{aligned}
\mathbb{E}[g(x)] &= \mathbb{E}[g(x_0)] + \mathbb{E}[g'(x_0)(x - x_0)] + \frac{1}{2}\mathbb{E}[g''(x_0)(x - x_0)^2] \\
&= \mathbb{E}[g(\mathbb{E}[X])] + \mathbb{E}[g'(\mathbb{E}[X])(x - \mathbb{E}[X])] + \frac{1}{2}\mathbb{E}[g''(\mathbb{E}[X])(x - \mathbb{E}[X])^2] \\
&= g(\mathbb{E}[X]) + g'(\mathbb{E}[X])\mathbb{E}[(x - \mathbb{E}[X])] + \frac{1}{2}g''(\mathbb{E}[X])\mathbb{E}[(x - \mathbb{E}[X])^2] \\
&= g(\mathbb{E}[X]) + g'(\mathbb{E}[X])(\mathbb{E}[x] - \mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\mathbb{E}[(x - \mathbb{E}[X])^2] \\
&= g(\mathbb{E}[X]) + g'(\mathbb{E}[X]) * 0 + \frac{1}{2}g''(\mathbb{E}[X])\mathbb{E}[(x - \mathbb{E}[X])^2] \\
&= g(\mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\mathbb{E}[(x^2 + \mathbb{E}[X]^2 - 2x\mathbb{E}[X]] \\
&= g(\mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\left(\mathbb{E}[x^2] + \mathbb{E}[\mathbb{E}[X]^2] - \mathbb{E}[2x\mathbb{E}[X]]\right) \\
&= g(\mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\left(\mathbb{E}[x^2] + \mathbb{E}[X]^2 - 2\mathbb{E}[X]^2]\right) \\
&= g(\mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\left(\mathbb{E}[x^2] - \mathbb{E}[X]^2]\right) \\
&= g(\mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\mathbb{V}[X].
\end{aligned}$$

Therefore, $\mathbb{E}[g(x)] \approx g(\mathbb{E}[X]) + \frac{1}{2}g''(\mathbb{E}[X])\mathbb{V}[X]$.

4.3 First Order Taylor Expansion to Find Variance

Let $X \sim f_x(x|\theta)$ and let $Y = g(x)$. What is the variance of Y? Using a similar approach, again making use of the Taylor Expansion approximation, we can answer this question. We will see where this comes in later on: let $x_0 = \mathbb{E}[X]$.

$$\begin{aligned}
\mathbb{V}[g(x)] &\approx \mathbb{V}[g(x_0) + g'(x_0)(X - x_0)] \\
&= \mathbb{V}[g(x_0)] + \mathbb{V}[g'(x_0)(X - x_0)] \\
&= 0 + \mathbb{V}[g'(x_0)(X - x_0)] \text{ since } g(x_0) \text{ is a constant} \\
&= g'(x_0)^2\mathbb{V}[(X - x_0)] \text{ since } g(x_0) \text{ is a constant} \\
&= g'(x_0)^2\left(\mathbb{V}[X] - \mathbb{V}[x_0]\right) \text{ since } g(x_0) \text{ is a constant}
\end{aligned}$$

$$= g'(x_0)^2\mathbb{V}[X] \text{ since } x_0 \text{ is a constant}$$
$$= g'(\mathbb{E}[X])^2\mathbb{V}[X] \text{ since } x_0 = \mathbb{E}[X].$$

So, $\mathbb{V}[X] \approx \mathbb{V}[X]g'(\mathbb{E}[X])$. $\square$

4.4 Second Order Taylor Expansion to Find Variance

Let $X \sim f_x(x|\theta)$ and let $Y = g(x)$. What is the variance of Y? We previously used the First Order Taylor Expansion. We can get an even closer approximation by using the second order. We will see where this comes in later on: let $x_0 = \mathbb{E}[X]$. Try this on your own before seeing my solutions.

$$\begin{aligned}
\mathbb{V}[g(x)] &\approx \mathbb{V}[g(x_0)] + \mathbb{V}[g'(x_0)(x - x_0)] + \mathbb{V}[\frac{1}{2}g''(x_0)(x - x_0)^2] \\
&= 0 + g'(x_0)^2\mathbb{V}[x - x_0] + \frac{g''(x_0)^2}{4}\mathbb{V}[(x - x_0)^2] \\
&= g'(x_0)^2\left(\mathbb{V}[x] - \mathbb{V}[x_0]\right) + \frac{g''(x_0)^2}{4}\mathbb{V}[x^2 + x_0^2 - 2(x)(x_0)] \\
&= g'(x_0)^2\mathbb{V}[x] + \frac{g''(x_0)^2}{4}\left(\mathbb{V}[x^2] + \mathbb{V}[x_0^2] - \mathbb{V}[2(x)(x_0)]\right) \\
&= g'(x_0)^2\mathbb{V}[x] + \frac{g''(x_0)^2}{4}\left(\mathbb{V}[x^2] + 0 - 4x_0^2\mathbb{V}[x]\right) \\
&= g'(x_0)^2\mathbb{V}[x] + \frac{g''(x_0)^2}{4}(1 - 4x_0^2)\mathbb{V}[x] \\
&= g'(x_0)^2\mathbb{V}[x] + \frac{g''(x_0)^2}{4}(1 - 4x_0^2)\mathbb{V}[x].
\end{aligned}$$

Now, setting $x_0 = \mathbb{E}[X]$, we see that

$$\mathbb{V}[g(x)] = g'(\mathbb{E}[X])^2\mathbb{V}[x] + \frac{g''(\mathbb{E}[X])^2}{4}(1 - 4\mathbb{E}[X]^2)\mathbb{V}[x].\square$$

5 Conclusion

We learned about the properties of the expectation and variance operators. We also learned how to find the expectation and variance of non-linear functions of our random variable using the delta-method, an application of Taylor Series Expansions. Some key notes:

- $\mathbb{E}[X] = \int_X x f_X dx$ with random variable X, support of X as S and PDF/PMF f_X
- $1 = \int_S f_X dx$ with support S of random variable X and PDF/PMF f_X
- $\mathbb{V}[X] = \int_S (X - \mathbb{E}[X])^2 f_X dx$ with support S of random variable X and PDF/PMF f_X
- $0 = \int_S (X - \mathbb{E}[X]) dx$

6 Appendix

6.1 Standard Distributions

Here is a list of distributions you are expected to know for this class along with their supports. To learn more about each, please see this document.

6.1.1 Normal Distribution - $x \in \mathbb{R}$

6.1.2 Gamma Distribution - $x \in (0, \infty)$

6.1.3 Exponential Distribution - $x \in (0, \infty)$

6.1.4 logNormal Distribution - $x \in (0, \infty)$

6.1.5 Chi-Squared with k degrees of freedom - $x \in (0, \infty)$

6.1.6 F Distribution with k_1, k_2 degrees of freedom - $x \in (0, \infty)$

6.1.7 T Distribution with k degrees of freedom - $x \in \mathbb{R}$

6.1.8 Beta Distribution - $x \in [0, 1]$

6.1.9 Bernoulli Distribution - $x \in \{0, 1\}$

6.1.10 Binomial Distribution - $x \in \{0, N\}$ for some natural number N

6.1.11 Geometric Distribution - $x \in \{0, 1, 2, ..., k\}$ for some natural number k

6.1.12 Poisson Distribution - $x \in \{0, 1, 2, ..., \infty\}$

6.1.13 Negative binomial Distribution - $x \in \{0, 1, 2, ..., \infty\}$