Mathematical Statistics I

Chapter 4: Expected Values

Jesse Wheeler

Outline

1. Discrete random variables

2. Continuous random variables

Discrete random variables

Introduction

- This material comes primarily from Rice (2007, Chapter 4).
- We will cover the ideas of expected value, variance, as well has higher-order moments.
- This includes topics such as conditional expectation, which is one of the fundamental ideas behind many branches of statistics and machine learning.
- For instance, most regression / prediction algorithms are built with the idea of minimizing some conditional expectation.

Expectation: Discrete random variables

Definition: Expectation of discrete random variables

Let X be a discrete random variable with pmf $p(x)$, which takes values in the space $\mathcal{X}$. The expected value of X is

$$E(X) = \sum_{x \in \mathcal{X}} x\, p(x),$$

provided that $\sum_{x \in \mathcal{X}} |x|\, p(x) < \infty$; otherwise, the expectation is not defined.

- This is not the most mathematically precise definition of expectation, but a more complete treatment of the topic is outside the scope of this course (See Resnick, 2019).

Expectation: Discrete random variables II

- The concept of the expected value parallels the notion of a *weighted average*.
- That is, we weight each possibility $x \in \mathcal{X}$ by their corresponding probability: $\sum_x x\, p(x)$.
- $E(X)$ is also referred to as the mean of X, and is typically denoted μ or μ_X.
- If the function p is thought of as a weight, then $E(X)$ is the center; that is, if we place the mass $p(x_i)$ at the points x_i, then the balancing point is $E(X)$.
- Like with the pmf and cdf, we often use subscripts to denote which probability law we are using for the expectation, it if is not clear: $E_X(X)$.

Expectation: Discrete random variables III

Roulette

A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 according to whether that event occurs. If X denotes your net gain, $X = 1$ with probability $18/38$ and $X = -1$ with probability $20/28$. The expected value of X is

$$E(X) = 1 \times \frac{18}{38} + (-1) \times \frac{20}{38} = -\frac{1}{19}.$$

- As you might imagine, the expected value coincides in the limit with the actual average loss per game, if you play many games (Chapter 5).

Expectation: Discrete random variables IV

- Most casino games have a negative expected value by design; you may win some money, but if a large number of games are played, the house will come out on top.

Geometric Random Variable

Suppose that items are produced in a plant are independently defective with probability p. If items are inspected one by one until a defective item is found, then how many items must be inspected on average?

Solution:

Expectation: Discrete random variables VI

Poisson Distribution

The Poisson(λ) distribution has pmf $p(k) = \frac{\lambda^k}{k!}e^{-\lambda}$, for all $k \geq 0$. Thus, if $X \sim$ Pois(λ), then what is $E[X]$?

Solution:

Continuous random variables

Expectation: Continuous random variables

Definition: Expectation of continuous random variables

Let X be a continuous random variable with pdf $f(x)$, which takes values in the space $\mathcal{X}$. The expected value of X is

$$E(X) = \int_{x \in \mathcal{X}} x f(x) \, dx.$$

provided that $\int_{x \in \mathcal{X}} x f(x) \, dx < \infty$, otherwise the expectation is undefined.

- As before, this is not the most mathematically precise definition of expectation, but a more complete treatment of the topic is outside the scope of this course (See Resnick, 2019).

Expectation: Continuous random variables II

- We can still think of $E(X)$ as the center of mass of the density.

Expectation: Continuous random variables III

Gamma Density

If X follows a gamma density with parameters α and λ, then the pdf of X is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0.$$

Find $E(X)$.

Solution.

Functions of random variables

- We are often interested in functions of random variables: $Y = g(X)$.
- Ideas that we have already covered enable us to calculate $E(Y)$.
- For instance, you could use the change-of-variables theorem to get the density of Y, then use the definition to calculate $E[Y]$.
- Fortunately, we don't have to do this. We can instead calculate $E[Y]$ by integrating (or summing) with respect to X:

$$E[g(X)] = \int_{x \in \mathcal{X}} g(x) f(x) \, dx.$$

- We will justify this for the discrete analog.

Functions of random variables II

Theorem 4.1: Expectation of transformed random variables

Suppose that X is a random variable and that $Y = g(X)$ for some function g. Then,

- If X is discrete with pmf $p(x)$:

$$E(Y) = \sum_x g(x)\,p(x),$$

  provided that $\sum_x |g(x)|p(x) < \infty$.

- If X is continuous with pdf $f(x)$:

$$E(Y) = \int_{-\infty}^{\infty} g(x) f(x)\,dx,$$

  provided that $\int |g(x)| f(x)\,dx < \infty$.

Functions of random variables: proof

Proof:

Functions of random variables: proof II

- The proof for the continuous case is similar, but does require a measure-theoretic approach to integration.
- One important thing to note is that $g\big(E(X)\big)$ is not usually equal to $E\big(g(x)\big)$.
- For example, let Z be a standard normal. We know that $E[Z] = 0$, because it's symmetric. However, $P\big(|Z| > 0\big) = 1$, thus we can readily deduce that $E\big[|Z|\big] \geq 0 = \big|E[Z]\big|$.
- An immediate consequence is that if for all non-negative random variables X that have finite expectation, if $g(x) \leq x$ for some function g, then $E[g(X)] \leq E[X]$.

Expected value of indicator functions

- An interesting example is indicator functions.
- For example, suppose that X is a random variable. Then $Y = 1[X \in A]$ for some $A \subset \mathcal{X}$ is a random variable.
- Example: Let X follow a standard normal distribution, and $A = [-1, 1]$. Then $Y = 1[X \in A]$ is defined as the random variables such that $Y(\omega) = 1$ if $X(\omega) \in A$, and $Y(\omega) = 0$ otherwise.

Expected value of indicator functions II

- Expectations of indicator variables are probabilities:

$$\begin{aligned} E(Y) &= E\big(1[X \in A]\big) \\ &= \int_{x \in \mathcal{X}} 1[X \in A]\, f(x)\, dx \\ &= \int_{x \in A} f(x)\, dx = P(X \in A). \end{aligned}$$

- This fact is useful for deriving some important inequalities.
- Let X be a continuous random variable with expectation $E(X)$. From our definition, this implies that $\int |x|\, f(x)\, dx < \infty$.

Expected value of indicator functions III

- Now suppose that for some random variable $Y = g(X)$ such that $|Y| \leq |X|$. Then, if Y has a pdf, we can deduce that $\int |y| \, f(x) \, dx < \infty$, and therefore $E[Y]$ exists.
- Now suppose that φ is a non-decreasing, non-negative function, and that for some $a \in \mathbb{R}$, $\varphi(a) > 0$. Then, for all $x \geq a$, $\varphi(x)/\varphi(a) \geq 1$.
- Define $Y = 1[X \geq a]$. Note that for all possible outcomes $\omega \in \Omega$,

$$Y = 1[X \geq a] \leq \varphi(X)/\varphi(a)1[X \geq a] \leq \varphi(X)/\varphi(a).$$

Expected value of indicator functions IV

- Taking expectations of both sides,

$$E\big(1[X \geq a]\big) = P(X \geq a) \leq \frac{E\big[\varphi(X)\big]}{\varphi(a)} = E\left[\varphi(X)/\varphi(a)\right].$$

- This inequality is known as Markov's (general) inequality, and is very useful for bounding the probability of particular events.
- Specifically, if $\varphi(x) = |x|^p$, with $p > 0$, then because $|X|$ is always positive, φ is non-negative, non-decreasing, and therefore

$$P(|X| \geq a) \leq \frac{E\big[|X|^p\big]}{a^p},$$

- If we restrict ourselves to the case where X is non-negative, we get the most standard version of the inequality:

$$P(X \geq a) \leq E(X)/a.$$

Markov's Inequality

- There are a

References and Acknowledgements

Resnick S (2019). *A probability path*. Springer.

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA.

- Compiled on August 25, 2025 using R version 4.5.1.

- We acknowledge students and instructors for previous versions of this course / slides.