Expectation of Random Variables

Saravanan Vijayakumaran
sarva@ee.iitb.ac.in

Department of Electrical Engineering
Indian Institute of Technology Bombay

February 12, 2014

Expectation of Discrete Random Variables

Definition

The expectation of a discrete random variable X with probability mass function f is defined to be

$$E(X) = \sum_{x:f(x)>0} xf(x)$$

whenever this sum is absolutely convergent. The expectation is also called the mean value or the expected value of the random variable.

Example

- Bernoulli random variable

  $\Omega = \{0, 1\}$

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

  where $0 \leq p \leq 1$

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

More Examples

- The probability mass function of a binomial random variable X with parameters n and p is

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{if } 0 \le k \le n$$

Its expected value is given by

$$E(X) = \sum_{k=0}^{n} kP[X = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

- The probability mass function of a Poisson random variable with parameter λ is given by

$$P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \ldots$$

Its expected value is given by

$$E(X) = \sum_{k=0}^{\infty} kP[X = k] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda$$

Why do we need absolute convergence?

- A discrete random variable can take a countable number of variables
- The definition of expectation involves a weighted sum of these values
- The order of the terms in the infinite sum is not specified in the definition
- The order of the terms can affect the value of the infinite sum
- Consider the following series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots$$

  Its sums to a value less than $\frac{5}{6}$
- Consider a rearrangement of the above series where two positive terms are followed by one negative term

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \cdots$$

  Since

$$\frac{1}{4k-3} + \frac{1}{4k-1} - \frac{1}{2k} > 0$$

  the rearranged series sums to a value greater than $\frac{5}{6}$

Why do we need absolute convergence?

- A series $\sum a_i$ is said to converge absolutely if the series $\sum |a_i|$ converges
- **Theorem:** If $\sum a_i$ is a series which converges absolutely, then every rearrangement of $\sum a_i$ converges, and they all converge to the same sum
- The previously considered series converges but does not converge absolutely
$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots$$
- Considering only absolutely convergent sums makes the expectation independent of the order of summation

Expectations of Functions of Discrete RVs

- If X has pmf f and $g : \mathbb{R} \to \mathbb{R}$, then

$$E(g(X)) = \sum_x g(x)f(x)$$

whenever this sum is absolutely convergent.

Example

- Suppose X takes values $-2, -1, 1, 3$ with probabilities $\frac{1}{4}, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}$ respectively.
- Consider $Y = X^2$. It takes values $1, 4, 9$ with probabilities $\frac{3}{8}, \frac{1}{4}, \frac{3}{8}$ respectively.

$$E(Y) = \sum_y yP(Y = y) = 1 \cdot \frac{3}{8} + 4 \cdot \frac{1}{4} + 9 \cdot \frac{3}{8} = \frac{19}{4}$$

Alternatively,

$$E(Y) = E(X^2) = \sum_x x^2 P(X = x) = 4 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 9 \cdot \frac{3}{8} = \frac{19}{4}$$

Expectation of Continuous Random Variables

Definition

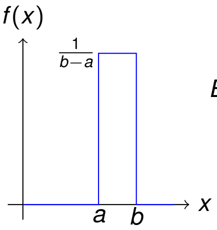
The expectation of a continuous random variable with density function f is given by

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

whenever this integral is finite.

Example (Uniform Random Variable)

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$



$$E(X) = \frac{a+b}{2}$$

Conditional Expectation

Definition

For discrete random variables, the conditional expectation of Y given $X = x$ is defined as

$$E(Y|X = x) = \sum_y y f_{Y|X}(y|x)$$

For continuous random variables, the conditional expectation of Y given X is given by

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\, dy$$

The conditional expectation is a function of the conditioning random variable i.e. $\psi(X) = E(Y|X)$

Example

For the following joint probability mass function, calculate $E(Y)$ and $E(Y|X)$.

Y/X	x_1	x_2	x_3
y_1	$\frac{1}{2}$	0	0
y_2	0	$\frac{1}{8}$	$\frac{1}{8}$
y_3	0	$\frac{1}{8}$	$\frac{1}{8}$

Law of Iterated Expectation

Theorem

The conditional expectation $E(Y|X)$ *satisfies*

$$E\left[E(Y|X)\right] = E(Y)$$

Example

A group of hens lay N eggs where N has a Poisson distribution with parameter λ. Each egg results in a healthy chick with probability p independently of the other eggs. Let K be the number of chicks. Find $E(K)$.

Some Properties of Expectation

- If $a, b \in \mathbb{R}$, then $E(aX + bY) = aE(X) + bE(Y)$
- If X and Y are independent, $E(XY) = E(X)E(Y)$
- X and Y are said to be uncorrelated if $E(XY) = E(X)E(Y)$
- Independent random variables are uncorrelated but uncorrelated random variables need not be independent

Example

Y and Z are independent random variables such that Z is equally likely to be 1 or -1 and Y is equally likely to be 1 or 2.

Let $X = YZ$. Then X and Y are uncorrelated but not independent.

Variance

- Quantifies the spread of a random variable
- If k is a positive integer, the kth moment m_k of X is defined to be

$$m_k = E(X^k)$$

- The kth central moment σ_k is

$$\sigma_k = E\left[(X - m_1)^k\right]$$

- The first moment is the same as the expectation $m_1 = E(X)$
- The second central moment $\sigma_2 = E[(X - m_1)^2]$ is called the variance
- The positive square root of the variance is called the standard deviation

$$\sigma = \sqrt{E\left[(X - m_1)^2\right]}$$

- Properties of Variance
  - $\text{var}(X) \geq 0$
  - $\text{var}(X) = E(X^2) - [E(X)]^2$
  - For $a, b \in \mathbb{R}$, $\text{var}(aX + b) = a^2 \text{var}(X)$
  - $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ if and only if X and Y are uncorrelated

Examples

- Variance of a binomial random variable X with parameters n and p is

$$\begin{aligned}\operatorname{var}(X) &= \sum_{k=0}^{n} k^2 P[X=k] - (np)^2 = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - n^2p^2 \\ &= np(1-p)\end{aligned}$$

- Variance of a Poisson random variable X with parameter λ is given by

$$\operatorname{var}(X) = \sum_{k=0}^{\infty} k^2 P[X=k] - \lambda^2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 = \lambda$$

- Variance of a uniform random variable X on $[a, b]$ is

$$\operatorname{var}(X) = \int_{-\infty}^{\infty} x^2 f_U(x)\, dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

Expectation via the Distribution Function

For a discrete random variable X taking values in $\{0, 1, 2, \ldots\}$, the expected value is given by

$$E[X] = \sum_{i=1}^{\infty} P(X \geq i)$$

Proof

$$\sum_{i=1}^{\infty} P(X \geq i) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} P(X = j) = \sum_{j=1}^{\infty} \sum_{i=1}^{j} P(X = j) = \sum_{j=1}^{\infty} jP(X = j) = E[X]$$

Example

Let $X_1, \ldots, X_m$ be m independent discrete random variables taking only non-negative integer values. Let all of them have the same probability mass function $P(X = n) = p_n$ for $n \geq 0$. What is the expected value of the minimum of $X_1, \ldots, X_m$?

Expectation via the Distribution Function

For a continuous random variable X taking only non-negative values, the expected value is given by

$$E[X] = \int_0^\infty P(X \geq x)\, dx$$

Proof

$$\begin{aligned} \int_0^\infty P(X \geq x)\, dx &= \int_0^\infty \int_x^\infty f_X(t)\, dt\, dx = \int_0^\infty \int_0^t f_X(t)\, dx\, dt \\ &= \int_0^\infty t f_X(t)\, dt = E[X] \end{aligned}$$

Probabilistic Inequalities

Markov's Inequality

If X is a non-negative random variable and $a > 0$, then

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof

We first claim that if $X \geq Y$, then $E(X) \geq E(Y)$.
Let Y be a random variable such that

$$Y = \begin{cases} a & \text{if } X \geq a, \\ 0 & \text{if } X < a. \end{cases}$$

Then $X \geq Y$ and $E(X) \geq E(Y) = aP(X \geq a) \implies P(X \geq a) \leq \frac{E(X)}{a}$.

Exercise

- Prove that if $E(X^2) = 0$ then $P(X = 0) = 1$.

Chebyshev's Inequality

Let X be a random variable and $a > 0$. Then $P\left(|X - E(X)| \geq a\right) \leq \frac{\text{var}(X)}{a^2}$.

Proof

Let $Y = (X - E(X))^2$.

$$P\left(|X - E(X)| \geq a\right) = P(Y \geq a^2) \leq \frac{E(Y)}{a^2} = \frac{\text{var}(X)}{a^2}.$$

Setting $a = k\sigma$ where $k > 0$ and $\sigma = \sqrt{\text{var}(X)}$, we get

$$P\left(|X - E(X)| \geq k\sigma\right) \leq \frac{1}{k^2}.$$

Exercises

- Suppose we have a coin with an unknown probability p of showing heads. We want to estimate p to within an accuracy of $\epsilon > 0$. How can we do it?
- Prove that $P(X = c) = 1 \iff \text{var}(X) = 0$.

Cauchy-Schwarz Inequality

For random variables X and Y, we have

$$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$$

Equality holds if and only if $P(X = cY) = 1$ for some constant c.

Proof

For any real k, we have $E[(kX + Y)^2] \geq 0$. This implies

$$k^2 E(X^2) + 2kE(XY) + E(Y^2) \geq 0$$

for all k. The above quadratic must have a non-positive discriminant.

$$[2E(XY)]^2 - 4E(X^2)E(Y^2) \leq 0.$$

Questions?