Lecture 6: Expectation

- Expectation and Variance
- The Poisson Distribution
- Continuous Random Variables

Introduction

– In the last lecture we learned about random variables.

– We also talked about some important discrete distributions.

– Today we will continue with how to calculate expectations for random variables.

– Last, we will revisit the definition of continuous random variables.

Probability Mass Function (PMF)

The **probability mass function (PMF)** of a discrete random variable is a function that maps k to $P(X=k)$. We have seen a few PMFs already. For instance,

$$f(k) = P(X=k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } 0 \leq k \leq n \\ 0, & \text{otherwise} \end{cases}$$

is the PMF of the Binomial(n, p). From now on, we will omit the "0 otherwise" case if it is reasonably clear from the context. The set $\mathcal{X} = \{k : P(X=k) > 0\}$ is called the **support** of the random variable. Clearly, we must have

$$\sum_{k \in \mathcal{X}} P(X=k) = 1$$

Cumulative Distribution Function (CDF)

The **cumulative distribution function (CDF)** of a discrete random variable is a function that maps k to $P(X \leq k)$. For instance, in the case of the Binomial(n, p):

$$F(k) = P(X \leq k) = \begin{cases} 0 & \text{if } k < 0 \\ \sum_{j=0}^{k} \binom{n}{j} p^k (1-p)^{n-k} & \text{if } 0 \leq k \leq n \\ 1 & \text{otherwise} \end{cases}$$

We will usually use the notation f for PMFs and F for CDFs.

Some useful properties of CDFs

Let $a, b \in \mathbb{R}$ be constants.

- $P(a < X \leq b) = F(b) - F(a)$
- $P(X > a) = 1 - F(a)$
- F is monotonically non-decreasing

Independent Random Variables

– Two discrete random variables X and Y are said to be independent if for all k_x, k_y

$$P(\{X = k_x\} \cap \{Y = k_y\}) = P(X = k_x)P(Y = k_y)$$

– Or equivalently in terms of the CDFs

$$P(\{X \leq k_x\} \cap \{Y \leq k_y\}) = P(X \leq k_x)P(Y \leq k_y)$$

– Yet another equivalent statement, now in terms of conditional probability, is that

$$P(X = k_x | Y = k_y) = P(X = k_x)$$

for all k_x and k_y.
– If X and Y are independent and g, h are functions, then $g(X)$ and $h(Y)$ are independent.

Expectation

– The **expected value** of a random variable is the average value of an infinite number of draws. It is often denoted by μ. Mathematically,

$$E[X] \equiv \mu \equiv \sum_{k \in \mathcal{X}} kP(X = k).$$

– This is a weighted average of the possible values of the random variable, where the weights are the probabilities that a value will occur.

– For example, the expected value of a fair die is 3.5. **Why?**

Properties of Expectations

Let $a, b, c \in \mathbb{R}$ be constants, and X and Y be random variables,

– **(Linearity)** Some functions are especially interesting. For example, the linear function $aX + bY + c$ has the useful property that $E[aX + bY + c] = aE[X] + bE[Y] + c$. This is true for all random variables X and Y, even if they are dependent

– If X and Y are independent, $E(XY) = E(X)E(Y)$

– **(Law of the Unconscious Statistician, LOTUS)** Let $h : \mathbb{R} \to \mathbb{R}$ be a function, then

$$E[h(X)] = \sum_{k \in \mathcal{X}} h(k)P(X = k)$$

– For example, the expected value of X^2 for a fair die is

$$\sum_{i=1}^{6} i^2 * \frac{1}{6} = 15.1667$$

Example

Suppose X and Y are independent random variables with PMFs given by the following table:

k	$P(X=k)$	$P(Y=k)$	$P(XY=k)$
0	0.3	0.3	0.51
1	0.7	0.5	0.35
2	0	0.2	0.14

The expectation of X is $E(X) = 0.3 \times 0 + 0.7 \times 1 = 0.7$ and the expectation of Y is $E(Y) = 0.5 + 0.2 \times 2 = 0.9$. Given that X and Y are independent, we have $E(XY) = E(X)E(Y) = 0.7 \times 0.9 = 0.63$. We could have found the expectation of XY using its PMF directly as $E(XY) = 1 \times 0.35 + 2 \times 0.14 = 0.63$. We can find $E(Y^2)$ using LOTUS: $E(Y^2) = 1^2 \times 0.5 + 2^2 \times 0.2 = 1.3$.

Variance

– The **variance** of a random variable is the expected value of $h(X) = (X - \mu)^2$ where μ is the expected value of X.

– The variance is often written as σ^2, and its units are the squares of the original units. It provides a measure of how spread out the sample is, since it is the average squared distance of an observation from μ. A little algebra shows:

$$\begin{aligned}
\text{Var}[X] &= E[(X-\mu)^2] \\
&= \sum_{k \in \mathcal{X}} (k-\mu)^2 P(X=k) \\
&= \sum_{k \in \mathcal{X}} (k^2 - 2k\mu + \mu^2) P(X=k) \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}$$

– For any random variable, the variance $V[X] \geq 0$

Properties of Variance

Let X, X_1, X_2 be random variables and $a, b \in \mathbb{R}$ be constants,

1. $V[aX + b] = a^2 V[X]$.
2. $V[X] = E[X^2] - E[X]^2$
3. If X_1 and X_2 are independent, $V(X_1 + X_2) = V(X_1) + V(X_2)$

   To see this, let $\mu_1 = E(X_1)$, $\mu_2 = E(X_2)$:

$$
\begin{aligned}
V(X_1 + X_2) &= E[(X_1 + X_2 - \mu_1 - \mu_2)^2] \\
&= E[(X_1 - \mu_1)^2] - 2E[X_1 - \mu_1]E[X_2 - \mu_2] + E[(X_2 - \mu_2)^2] \\
&= V(X_1) + V(X_2)
\end{aligned}
$$

Note that X_1 and X_2 does not have to be independent. Independence is sufficient but not necessary for the variance of the sum to equal the sum of the variances.

Standard Deviation

– To measure how spread out a distribution is, we mostly use the **standard deviation** (or σ). This is the square root of the variance.

– The variance of the result of a roll of a fair die is $E[X^2] - E[X]^2 = 15.1667 - (3.5)^2 = 2.9167$. Its standard deviation is $\sqrt{2.9167} = 1.7078$.

– One can show that the mean, variance and standard deviation of the $\text{Bin}(n, p)$ distribution are np, $np(1-p)$ and $\sqrt{np(1-p)}$, respectively.

Bernoulli and Binomial Distributions

- **Bernoulli:** The expectation of $X \sim \text{Bernoulli}(p)$ random variable is simply $E(X) = 0 \times (1-p) + 1 \times p = p$. Note that $X^2 = X$, so $E(X^2) = p$. Therefore, $V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$
- **Binomial:** Let $Y \sim \text{Binomial}(n, p)$. Then $Y = X_1 + X_2 + \ldots + X_n$, where $X_i \sim \text{Bernoulli}(p)$, and the X_i are independent. Therefore, $E(Y) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$, similarly, $V(Y) = V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n) = np(1-p)$.
- Note that we didn't use the independence of the X_i for deriving $E(Y)$, but we did for $V(Y)$
- Some textbook apply the definition of expectation and variance to the PMF of the Binomial and compute

$$E(Y) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}, \quad E(Y^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k},$$

using the properties of combinations and sums, which is way more tedious!

Expected Value and its Variability

– A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 100 Americans, how many would you expect to be obese?

Easy enough, for Binomial distribution, $\mu = np = 100 \times 0.262 = 26.2$.

– But this doesn't mean in every random sample of 100 people exactly 26.2 will be obese. In fact, that's not even possible. In some samples this value will be less, and in others more. How much would we expect this value of vary?

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

– We would expect 26.2 out of 100 randomly sampled Americans to be obese, with a standard deviation of 4.4.

Poisson Distribution

– Another important distribution is the **Poisson distribution**.

– The Poisson distribution gives the probability of exactly x occurrences in a fixed period of time or a fixed area. For example,

- The number of phone calls received by a call center per hour
- The number of taxis passing a particular street corner per hour
- The number of accidents on the highway in one month.
- The number of thefts on Duke campus between 2001 and 2016.

– The Poisson is used when a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event.

Poisson Distribution

– Suppose the mean number of events is λ. Then X has the Poisson distribution Pois(λ) and the PMF of $X \sim \text{Poisson}(\lambda)$ is:

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

– The expectation of a Poisson(λ) is a very fun calculus exercise.

$$E(X) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

We won't derive the variance here, but it can be shown that $V(X) = \lambda$, too.

WWII V2 Rocket Attacks

– *Example 5:* Military analysts divided London into 576 areas and studied 500 V-2 attacks. So the average number of attacks per block was 500/576 = 0.868.

– What is the probability that Buckingham Palace was undamaged?

$$\mathbb{P}[X = k] = \frac{0.868^0}{0!} e^{-0.868} = 0.419.$$

– What is the probability of at most one hit on Buckingham Palace?

$$\mathbb{P}[\text{at most one hit on BP }] = \frac{0.868^0}{0!} e^{-0.868} + \frac{0.868^1}{1!} e^{-0.868} = 0.784.$$

Examples

– *Example 6 (D.S. 5.4.1 – to be done in class):* A store owner believes that customers arrive at his store at a rate of 4.5 customers per hour on average. What is the probability that 5 customers would arrive in the next hour?

– There is an interesting relationship between the binomial and Poisson distributions. Turns out the $\text{Bin}(n, p)$ can be approximated by $\text{Po}(\lambda = np)$ as $n \to \infty$ and $p \to 0$. The larger the n and the smaller the p, the better the approximation.

Properties

– **Law of Rare Events:** Let $Y \sim \text{Binomial}(n, p)$ and $X \sim \text{Poisson}(np)$. If $n \to \infty$, $p \to 0$ but n goes to infinity at roughly the same rate as p goes to zero, so that $np \to \lambda$ (λ is a constant). Then

$$P(Y = k) \to P(X = k) \quad \text{as } n \to \infty$$

In practice, this means that if p is small, n is big, and np is not too close to 0 or too big, we can approximate $Y \sim \text{Binomial}(n, p)$ with $X \sim \text{Poisson}(np)$.

– **Additivity:** If $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$, $\cdots$, $X_n \sim \text{Poisson}(\lambda_n)$, and X_1, X_2, $\cdots$, X_n are independent, then $S = X_1 + X_2 + \cdots + X_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$.

Example: Slip and Fall

Let Y be the number of days in a year that I slip and fall just by walking around. I'm going to assume $Y \sim \text{Binomial}(365, 0.01)$. Let's approximate this with $X \sim \text{Poisson}(3.65)$. The probability that I don't slip and fall in a year (not even once!) is $P(Y = 0) = 0.99^{365} \approx 0.026$, and the Poisson approximation is $P(X = 0) = e^{-3.65} \approx 0.026$, which is pretty good.

Try to compute the Binomial probability with your calculator; it will probably give an error because "n choose k" factor will be too big, but you'll be able to compute the Poisson probability easily. Just for fun, the probability that I slip and fall at least 10 times a year is $P(Y \geq 10) \approx 0.004$ (and we can get an identical figure up to 3 decimals with the Poisson approximation)

Continuous Random Variables

– Recall that a random variable X has a **continuous distribution** if X can take on infinite values. For continuous distributions, specific values have probability zero. One can only assign probabilities to intervals

– For example, there is a positive probability of tossing exactly five heads in 10 flips. But there is zero probability of finding someone who is exactly five feet tall, when height is measured to an infinite number of decimal places.

– To define probabilities for intervals, we use a **density function**. A density function is any function $f(x)$ such that

- $f(x)$ is non-negative
- $f(x)$ integrates to 1.

Then

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f(x)\, dx.$$

Continuous Random Variables (Cont'd)

This definition of the density function ensures the probability axioms are satisfied:

1. All probabilities are between 0 and 1, inclusive,
2. $\mathbb{P}[-\infty < X < \infty] = 1$, and
3. If A and B are disjoint intervals, then

$$\mathbb{P}[X \in A \text{ or } X \in B] = \mathbb{P}[X \in A] + \mathbb{P}[X \in B].$$

It turns out to be useful to define the **cumulative distribution function** (or cdf) as

$$F(x) = \mathbb{P}[X \leq x] = \int_{-\infty}^{x} f(y)\, dy.$$

Then $f(x) = \dfrac{\mathrm{d}F(x)}{\mathrm{d}x}$, and

$$\mathbb{P}[a \leq X \leq b] = \int_{a}^{b} f(x)\, dx = F(b) - F(a).$$

Continuous Random Variables (Cont'd)

The expected value of a continuous random variable is

$$E[X] \equiv \mu \equiv \int_{-\infty}^{\infty} x f(x)\, dx.$$

Clearly, this is similar to the definition of expected value for a discrete random variable:

$$E[X] \equiv \mu \equiv \sum_{k \in \mathcal{X}} k p(X = k).$$

Analogously, the expected value of a function $h(X)$ of a continuous random variable is

$$E[h(X)] \equiv \int_{-\infty}^{\infty} h(x) f(x)\, dx.$$

As before, $E[aX + b] = aE[X] + b$. And $V[X] = E[X^2] - (E[X])^2$.

Expectation of a Continuous Random Variable

Example 1 (D.S. 4.1.6): An appliance has a maximum lifetime of one year. The time X until it fails is a random variable whose p.d.f is:

$$f(x) = \begin{cases} 2x & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = 5X^4$. Then,

$$E[X] = \int_{-\infty}^{\infty} x(2x)\mathrm{d}x = \int_0^1 2x^2 \mathrm{d}x = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$$

$$E[Y] = E[5X^4] = \int_{-\infty}^{\infty} 5x^4(2x)\mathrm{d}x = \int_0^1 10x^5 \mathrm{d}x = \frac{10x^6}{6}\Big|_0^1 = \frac{5}{3}$$

Expectation of a Continuous Random Variable Cont'd

Example 2: Suppose that a random variable X has pdf $f(x) = c$ for some constant c, where $1 \leq x \leq 3$. Can we find its expected value and variance such that it doesn't involve c? **Of course!**

Since the pdf must integrate to 1, we know how to find c. That is,

$$1 = \int_1^3 c \, \mathrm{d}x = cx\Big|_1^3 = 3c - c = 2c \Rightarrow c = \frac{1}{2}$$

$$\text{Then,} \quad E[X] = \int_1^3 \frac{x}{2} \, \mathrm{d}x = \frac{x^2}{4}\Big|_1^3 = \frac{9}{4} - \frac{1}{4} = 2$$

$$E[X^2] = \int_1^3 \frac{x^2}{2} \, \mathrm{d}x = \frac{x^3}{6}\Big|_1^3 = \frac{27}{6} - \frac{1}{6} = \frac{26}{6} = 4.333$$

$$\Rightarrow V[X] = E[X^2] - (E[X])^2 = 4.333 - 2^2 = 0.333$$

It turns out that this is another well known distribution. A random variable is said to have a uniform distribution (continous) over its support $a \leq x \leq b$ if $f(x) = c$ for some constant c. This is denoted $X \sim \text{Uniform}(a, b)$.

Properties of Expectation

Now lets review two more interesting properties of expectations.

1. If $X_1, X_2, \ldots, X_n$ are n random variables such that each expectation is finite and well-defined, then

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

2. If $X_1, X_2, \ldots, X_n$ are n "independent" random variables such that each expectation is finite and well-defined, then

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E[X_i]$$

Recap

We discussed the following:

- Calculating expectation and variance
- The Poisson distribution
- Continuous random variables

Suggested reading:

- D.S. Sec. 3.2, 3.3, 4.1, 4.2, 4.3, 5.4
- OpenIntro3: Sec. 2.4, 2.5, 3.5.2