Week 4

Statistics 251

Zhongjian Wang

Department of Statistics
The University of Chicago

Where are we?

Bernoulli and binomial random variables

Poisson random variables

Geometric random variables

Negative binomial random variables

Binomial random variables

Toss a fair coin n times. What is the probability of k heads?

What if the coin has probability p of coming up heads?

Binomial random variables

A **Bernoulli** random variable with parameter $p \in [0,1]$ has value

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

A **binomial** random variable with parameters $n \geq 1, p \in [0,1]$ has value

$$X = k \text{ with probability } \binom{n}{k} p^k (1-p)^{n-k}.$$

If $X_i \sim \text{Bernoulli}(p)$ for $i = 1, \ldots, n$, then

$$X_1 + X_2 + \cdots + X_n \sim \text{Binomial}(n, p).$$

Binomial random variables

- Flip a coin 10 times: number of H is Binomial$(10, \frac{1}{2})$.
- Survey 100 people chosen at random in Chicago, and ask a yes/no question—this type of experiment is often modeled as Binomial (even though technically the sampling is usually without replacement)
- **Not** binomial: draw 10 cards from a standard deck and count the number of red cards.
- **Not** binomial: roll a dice until the first time you get a 6. How many rolls did it take?

Examples

If a room contains n people, what is the probability that exactly k of them were born on Monday?

If $n = 100$, what is the probability that at most 98 of them were born on Monday?

Expectation

If $X \sim \text{Binomial}(n, p)$, what is $\mathbb{E}[X]$? We may compute

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k\mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&= np.
\end{aligned}$$

Expectation (alternate approach)

If $X \sim \text{Binomial}(n, p)$, what is $\mathbb{E}[X]$?

Recall $X = X_1 + \cdots + X_n$, where $X_i \sim \text{Bernoulli}(p)$.

By linearity of expectation, we have

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1 + \cdots + X_n] \\ &= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] \\ &= np,\end{aligned}$$

where we note that $\mathbb{E}[X_i] = p$.

Variance

If $X \sim \text{Binomial}(n, p)$, what is $\text{Var}(X)$? We may compute

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{n} k^2 \mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} k \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&\quad + n(n-1)p^2 \sum_{k=0}^{n} \frac{(n-1)!}{(n-k)!(k-2)!} p^{k-2} (1-p)^{n-k}
\end{aligned}$$

Variance

If $X \sim \text{Binomial}(n, p)$, what is $\text{Var}(X)$? We may compute

$$
\begin{aligned}
\mathbb{E}[X^2] &= np \sum_{k=0}^{n} k \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&\qquad + n(n-1)p^2 \sum_{k=0}^{n} \frac{(n-2)!}{(n-k)!(k-2)!} p^{k-2} (1-p)^{n-k} \\
&= np + n(n-1)p^2 \\
&= np + n^2p^2 - np^2.
\end{aligned}
$$

In addition, we have $\mathbb{E}[X]^2 = n^2p^2$, so

$$
\text{Var}(X) = np + n^2p^2 - np^2 - n^2p^2 = np(1-p).
$$

Variance (alternate approach)

If $X \sim \text{Binomial}(n, p)$, what is $\text{Var}[X]$?

Recall $X = X_1 + \cdots + X_n$, where $X_i \sim \text{Bernoulli}(p)$.

We may compute

$$\mathbb{E}[X^2] = \mathbb{E}\left[\sum_{i,j=1}^{n} X_i X_j\right] = \sum_{i,j=1}^{n} \mathbb{E}[X_i X_j]$$
$$= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{i<j} \mathbb{E}[X_i X_j] = np + n(n-1)p^2.$$

As before, this means

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = np(1-p).$$

Problems

An airplane seats 198, but the airline has sold 200 tickets. Each person independently has a 0.05 chance of not showing up for the flight. What is the probability that more than 198 people will show up for the flight?

Problems

In a 100 person senate, 40 people always vote for the Republicans, 40 people always vote for the Democrats, and 20 people toss a coin to decide which way to vote. What is the probability that a given vote is tied?

Where are we?

Bernoulli and binomial random variables

Poisson random variables

Geometric random variables

Negative binomial random variables

Motivation

Consider the following random variables:

- The number of plane crashes in a year.
- The number of calls to a call center in an hour.
- The number of goals during a 90 minute soccer match.
- The number of raindrops which hit an umbrella in a minute.

In all of these examples, we can divide the time interval into small **increments** and add up the count during each interval. The count in one interval is **approximately independent** of other intervals.

Motivation

Divide a time interval into N **increments** and **assume** each increment has at most 1 event, with probability $\frac{\lambda}{N}$.

The total number of events is

$$\text{Binomial}(N, \frac{\lambda}{N}).$$

What happens as $N \to \infty$?

Reminder on e

Recall that

$$e = \lim_{n\to\infty} (1 + 1/n)^n$$

and

$$e^\lambda = \lim_{n\to\infty} (1 + \lambda/n)^n$$

and

$$e^{-\lambda} = \lim_{n\to\infty} (1 - \lambda/n)^n.$$

Poisson random variables

Let $\lambda > 0$ be a number, and let n be huge (e.g. 10^6).

Suppose I take n tosses of a coin that comes up heads with probability $p = \frac{\lambda}{n}$.

How many heads do I expect?

- Answer: $n \cdot p = \lambda$.

What is the probability I get k heads?

- Binomial(n, p), so

$$\begin{aligned}\mathbb{P}(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{\lambda^k}{n^k} (1 - \lambda/n)^{n-k} \\ &\approx \frac{\lambda^k}{k!} e^{-\lambda}.\end{aligned}$$

Poisson random variables

A **Poisson** random variable with parameter λ has

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Poisson normalization

A **Poisson** random variable with parameter λ has

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

How can we show $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$?

We have that

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

where we note by Taylor expansion that

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$$

Poisson expectation

A **Poisson** random variable with parameter λ has

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

What is $\mathbb{E}[X]$? ($\approx$ Binomial$(n, \lambda/n)$, so expect λ)

We may compute that

$$\begin{aligned}
\mathbb{E}[X] = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda.
\end{aligned}$$

Poisson variance

A **Poisson** random variable with parameter λ has

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

What is $\mathrm{Var}[X]$?

We may compute that

$$\begin{aligned}
\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \mathbb{P}(X = k) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
= \sum_{k=0}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda} &= \lambda \sum_{k=0}^{\infty} (1 + (k-1)) \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\
&= \lambda \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} + \lambda^2 \sum_{k=0}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2 + \lambda.
\end{aligned}$$

This means $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda$.

Problems

A country has an average of 2 plane crashes per year. Is it reasonable to assume the number of crashes in a year is Poisson with parameter 2?

Under this assumption, what is the probability of exactly 2 crashes? What about 0?

Problems

A city has an average of 5 major earthquakes a century. If the number of earthquakes is Poisson, what is the probability there is at least an earthquake in a given decade?

An online poker site deals 1 million poker hands per day. Approximate the probability that there are exactly 2 royal flush hands in a given day. (probability of royal flush: 1.54×10^{-7})

Where are we?

Bernoulli and binomial random variables

Poisson random variables

Geometric random variables

Negative binomial random variables

Geometric random variables

Consider independent tosses of a coin which is heads with probability p. A **geometric** random variable with parameter p is the toss number of the first appearance of heads.

We have that

$$\mathbb{P}(X = k) = p(1-p)^{k-1}.$$

Properties of geometric random variables

For a geometric random variable, we have

$$\mathbb{P}(X = k) = p(1-p)^{k-1}.$$

What is $\mathbb{E}[X]$?

By definition, we have

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} kp(1-p)^{k-1}.$$

Notice that

$$\begin{aligned}
\mathbb{E}[X] - 1 &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} - \sum_{k=1}^{\infty} p(1-p)^{k-1} \\
&= \sum_{k=1}^{\infty} (k-1)p(1-p)^{k-1} = \sum_{j=1}^{\infty} jp(1-p)^{j} = (1-p)\mathbb{E}[X].
\end{aligned}$$

We can solve to find $\mathbb{E}[X] = 1/p$.

Properties of geometric random variables

For a geometric random variable, we have

$$\mathbb{P}(X = k) = p(1-p)^{k-1}.$$

What is $\mathrm{Var}(X)$?

Notice that

$$\mathbb{E}[(X-1)^2] = \sum_{k=1}^{\infty}(k-1)^2 p(1-p)^{k-1} = \sum_{j=1}^{\infty} j^2 p(1-p)^j = (1-p)\mathbb{E}[X^2].$$

We find that

$$p\mathbb{E}[X^2] = 2\mathbb{E}[X] - 1 = \frac{2}{p} - 1 \implies \mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}.$$

We conclude that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1-p}{p}$.

Memoryless property

The geometric distribution is **memoryless**: given that you have flipped a coin 100 times with no successes so far, the distribution of the additional number of flips needed is unchanged.

Does this apply for...

- Waiting for the bus?
- Flipping a coin you picked up off the ground?
- Applying for jobs?

Where are we?

Bernoulli and binomial random variables

Poisson random variables

Geometric random variables

Negative binomial random variables

Negative binomial random variables

Consider independent tosses of a coin which is heads with probability p. A **negative binomial** random variable with parameters r and p is the toss number of the rth head.

If $r = 1$, get a geometric random variable with parameter p.

To get the rth head on the kth toss, we need exactly $r - 1$ heads among the first $k - 1$ tosses. There are $\binom{k-1}{r-1}$ ways to choose these tosses, implying

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} p.$$

Negative binomial properties

Consider independent tosses of a coin which is heads with probability p. A **negative binomial** random variable with parameters r and p is the toss number of the rth head.

What is $\mathbb{E}[X]$? Notice that

$$X = X_1 + X_2 + \cdots + X_r,$$

where X_i is geometric with parameter p. This implies that

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_r] = \frac{r}{p}.$$

Negative binomial properties

Consider independent tosses of a coin which is heads with probability p. A **negative binomial** random variable with parameters r and p is the toss number of the rth head.

What is $\mathrm{Var}(X)$? Notice that $X = X_1 + X_2 + \cdots + X_r$ where X_i is geometric with parameter p. This implies that

$$\begin{aligned}\mathbb{E}[X^2] = \mathbb{E}\left[\sum_{i,j=1}^{r} X_i X_j\right] &= \sum_{i=1}^{r} \mathbb{E}[X_i^2] + 2 \sum_{1 \le i < j \le r} \mathbb{E}[X_i X_j] \\ &= r\frac{2-p}{p^2} + r(r-1)\frac{1}{p^2}.\end{aligned}$$

We find

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = r\frac{2-p}{p^2} - \frac{r}{p^2} = \frac{r(1-p)}{p^2}.$$

Summary

Consider a sequence of independent tosses of a coin which is heads with probability p.

- each toss is Bernoulli(p)
- the number of heads in the first n tosses is Binomial(n, p)
- the toss number of the first head is Geometric(p)
- the toss number of the r^{th} head is NegativeBinomial(r, p)

Problems

A dog barks with probability 0.01 each minute.

How many times do we expect the dog to bark between noon and midnight?

What is the probability the dog is quiet between noon and 2pm and barks at exactly 2pm?

Problems

A dog barks with probability 0.01 each minute.

What is the probability the dog is quiet between noon and 2pm?

What is the probability the fifth bark since noon is at midnight?

Problems

A dog barks with probability 0.01 each minute.

How many minutes do I expect to wait until the fifth bark?

Approximate the probability there are exactly 5 barks between noon and midnight.

Problems

Consider the following game:
At each round, you roll a red die & a blue die.
If the red die is even, you win a prize, otherwise you win nothing.
If the blue die is a 1, then you stop playing, otherwise you continue.

1. What is the distribution of the # of rounds you play?
2. What is the distribution of the # of times you win?