Recitation/Discussion 5

Binomial, Geometric, Negative Binomial, Poisson, and Continuous Distributions

accompaniment to MATH 170E: Introduction to Probability and Statistics 1: Probability at the University of California, Los Angeles

Allen Lin

In this recitation, we will review discrete random variables. Then, we will introduce continuous random variables.

Exit Survey

Please fill out the exit survey.

This is so I can get feedback and know what to focus on next time!

Discrete Distributions

The following table summarizes the data.

X	$P(X = k)$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
$X \sim \mathrm{Bin}(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
$X \sim \mathrm{Geom}(p)$	$(1-p)^{k-1} p$	$1/p$	$(1-p)/p^2$
$X \sim \mathrm{NB}(r, p)$	$\binom{k-1}{r-1} (1-p)^{k-r} p^r$	r/p	$r(1-p)/p^2$
$X \sim \mathrm{Pois}(\lambda)$	$\lambda^k e^{-\lambda}/k!$	λ	λ

Binomial Distributions

Definition 1. The *binomial distribution* with parameters n and p is the discrete probability distribution of the number of successes of n independent trials, with each trial resulting in a success with probability p. If X is a random variable that follows the binomial distribution with n trials with probability p, we write $X \sim \mathrm{Bin}(n, p)$. Each experiment is known as a *Bernoulli trial*.

Proposition 2. *Let $X \sim \mathrm{Bin}(n, p)$. Then X follows the pmf*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

An easy way to remember this: there are k successes out of n experiments, of which there are $\binom{n}{k}$ such arrangements.

for $k \in \{0, \ldots, n\}$.

Proof. Let $k \in \{0, \ldots, n\}$. For a fixed choice of k successes out of n experiments, there are $n - k$ failures, with a resulting probability of $p^k (1-p)^{n-k}$. There are $\binom{n}{k}$ ways to choose the positions of k successes, so the total probability is $\binom{n}{k} p^k (1-p)^{n-k}$. ■

Proposition 3. *If $X \sim \text{Bin}(n, p)$, then $\mathbb{E}[X] = np$.*

This is an excellent application of indicator random variables.

Proof. Let $\mathbb{I}_i$ be the indicator variable for success on the ith experiment. Then $X = \mathbb{I}_1 + \cdots + \mathbb{I}_n$, so by linearity of expectation

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{I}_1 + \cdots + \mathbb{I}_n] = \mathbb{E}[\mathbb{I}_1] + \cdots + \mathbb{E}[\mathbb{I}_n] = \underbrace{p + \cdots + p}_{n \text{ times}} = np,$$

Again, expressing a sum of variables as a sum of indicator variables will be a recurring idea.

which finishes the proof. ■

Proposition 4. *If $X \sim \text{Bin}(n, p)$, then $\text{Var}[X] = np(1-p)$.*

Proof 1. Let $\mathbb{I}_i$ be the indicator variable for success on the ith experiment. Then $X = \mathbb{I}_1 + \cdots + \mathbb{I}_n$. Since $\mathbb{I}_1, \ldots, \mathbb{I}_n$ **are independent**, we can write

This is the slicker proof. It turns out that variance is also linear **under independence**, which by assumption we have.

$$\text{Var}(X) = \text{Var}(\mathbb{I}_1 + \cdots + \mathbb{I}_n) = \text{Var}(\mathbb{I}_1) + \cdots + \text{Var}(\mathbb{I}_n) = np(1-p)$$

since $\text{Var}(\mathbb{I}_i) = \mathbb{E}\left[\mathbb{I}_i^2\right] - \mathbb{E}[\mathbb{I}_i]^2 = \mathbb{E}\left[\mathbb{I}_i\right] - \mathbb{E}[\mathbb{I}_i]^2 = p - p^2 = p(1-p)$. ■

Proof 2. We just expand $X^2 = (\mathbb{I}_1 + \cdots + \mathbb{I}_n)^2$ to get

I like this proof too.

$$\mathbb{E}\left[X^2\right] = \mathbb{E}[(\mathbb{I}_1 + \cdots + \mathbb{I}_n)^2] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_i^2 + \sum_{i \neq j} \mathbb{I}_i \mathbb{I}_j\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}[\mathbb{I}_i] + \sum_{i \neq j} \mathbb{E}[\mathbb{I}_i \mathbb{I}_j] = np + n(n-1)p^2$$

Linearity of expectation, $\mathbb{I}_i^2 = \mathbb{I}_i$, $n(n-1)$ pairs $i \neq j$ (why?), and $\mathbb{E}[\mathbb{I}_i \mathbb{I}_j] = p^2$ (why?)

so $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X^2] = np + n(n-1)p^2 - (np)^2 = np(1-p)$. ■

Proof 3. We recall that $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X] = \mathbb{E}\left[X^2\right] - (np)^2$, so we need only calculate $\mathbb{E}\left[X^2\right]$. This is just a computation:

You should verify each step of the computation.

$$\mathbb{E}\left[X^2\right] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}$$ By definition

$$= \sum_{k=0}^{n} kn \binom{n-1}{k-1} p^k (1-p)^{n-k}$$ $k^2\binom{n}{k} = kn\binom{n-1}{k-1}$

$$= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$ $k = 0$ term is 0 and reindexing

$$= np \sum_{j=0}^{m} (j+1) \binom{m}{j} p^j (1-p)^{m-j}$$ $j = k-1$ and $m = n-1$

$$= np \left(\sum_{j=0}^{m} j \binom{m}{j} p^j (1-p)^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \right)$$ Splitting sum into two

$$= np \left(p \sum_{j=1}^{m} j \binom{m}{j} p^{j-1} (1-p)^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \right)$$ $j = 0$ term is 0 and reindexing

$$= np \left(p \sum_{j=1}^{m} m \binom{m-1}{j-1} p^{j-1} (1-p)^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j} \right)$$ $j\binom{m}{j} = m\binom{m-1}{j-1}$

$$= np \left(mp(p + (1-p))^{m-1} + (p + (1-p))^m \right) = n^2 p^2 + np(1-p),$$ Binomial theorem

so $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = n^2 p^2 + np(1-p) - (np)^2 = np(1-p)$. ■

Geometric Distributions

Definition 5. The *geometric distribution* with parameter p is the discrete probability distribution of the number of trials needed for one success, with each trial resulting in a success with probability p. If X is a random variable that follows the geometric distribution with probability p, we write $X \sim \text{Geom}(p)$.

Proposition 6. *Let $X \sim \text{Geom}(p)$. Then X follows the pmf*

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

An easy way to remember this: $k-1$ failures before success.

for $k \in \mathbb{N}$.

Proof. Let $k \in \mathbb{N}$. If the first success is on the kth experiment, then the first $k-1$ are failures. The latter happens with probability $(1-p)^{k-1}$, and the kth experiment occurs with success with probability p. The total probability is thus $(1-p)^{k-1}p$. ■

Proposition 7. *If $X \sim \text{Geom}(p)$, then $\mathbb{E}[X] = 1/p$.*

This should make sense from an intuitive standpoint: You should expect two coin flips for a head, or six dice rolls for a 5.

Proof 1. The first trial either succeeds with probability p or fails with probability $1-p$. If it fails, the remaining expected number of trials until a success is identical to the original expected number (if we just disregard the first trial) since all the trials are independent. Therefore,

This is the slicker proof.

$$\mathbb{E}[X] = p + (1-p)(1 + \mathbb{E}[X]),$$

which gives $\mathbb{E}[X] = 1/p$. ■

Proof 2. Recall the sum of the geometric series is

$$a + \sum_{k=1}^{\infty} ar^k = \sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$$

for $|r| < 1$. Differentiating with respect to r gives

This is justified by the uniform convergence of power series.

$$\sum_{k=1}^{\infty} kar^{k-1} = \frac{a}{(1-r)^2}$$

for $|r| < 1$. We just calculate the expected value

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} kP(X = k) \\
&= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= \frac{p}{(1-(1-p))^2} \\
&= \frac{1}{p},
\end{aligned}$$

Here $a = 1$ and $r = 1-p$.

as required. ■

Proposition 8. *If $X \sim \text{Geom}(p)$, then $\text{Var}(X) = (1-p)/p^2$.*

Proof. Recall that

$$\sum_{k=1}^{\infty} kar^{k-1} = \frac{a}{(1-r)^2}$$

for $|r| < 1$. Therefore,

$$\sum_{k=1}^{\infty} kar^{k} = \frac{ar}{(1-r)^2}$$

Differentiating with respect to r gives

$$\sum_{k=1}^{\infty} k^2 ar^{k-1} = \frac{a(1-r)^2 + 2ar(1-r)}{(1-r)^4} = \frac{a(1-r) + 2ar}{(1-r)^3} = \frac{a(1+r)}{(1-r)^3}$$

for $|r| < 1$. We just calculate

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= \sum_{k=1}^{\infty} k^2 P(X = k) \\
&= \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p \\
&= \frac{p(1 + (1-p))}{(1-(1-p))^3} \\
&= \frac{2-p}{p^2}.
\end{aligned}$$

Therefore,

$$\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2},$$

as required. ■

Negative Binomial Distributions

Definition 9. The *negative binomial distribution* with parameters r and p is the discrete probability distribution of the number of trials needed for r successes, with each trial resulting in a success with probability p. If X is a random variable that follows the negative binomial distribution for r successes with probability p, we write $X \sim \text{NB}(r, p)$.

Also known as a *Pascal distribution*.

Note that $\text{Geom}(p) = \text{NB}(1, p)$.

Proposition 10. *Let $X \sim \text{NB}(r, p)$. Then X follows the pmf*

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

for $k \geq r$.

An easy way to remember this: The kth trial is a success, and the remaining $r-1$ successes are out of the last $k-1$ trials.

Proof. Let $k \geq r$. If the rth success is on the kth experiment, then the previous $r-1$ successes occurred in the previous $k-1$ experiments. There are $\binom{k-1}{r-1}$ such arrangements, each with probability $p^{r-1}(1-p)^{(k-1)-(r-1)} = p^{r-1}(1-p)^{k-r}$. Accounting for the rth success which happens with probability p, the total probability is thus $\binom{k-1}{r-1}p^r(1-p)^{k-r}$. ■

Proposition 11. *If $X \sim \mathrm{NB}(r,p)$, then $\mathbb{E}[X] = r/p$.*

Note that this is r times as much as the expectation of a geometric random variable. This might suggest a linearity of expectation proof where we have r random variables, one for each success.

Proof. For each $i \in \{1,\ldots,r\}$ let Y_i be the random variable denoting the number of trials needed to obtain the ith success after the $(i-1)$th success. Then

$$X = Y_1 + \cdots + Y_r$$

counts the number of trials needed for r successes, so

This is a common trick.

$$\mathbb{E}[X] = \mathbb{E}[Y_1 + \cdots + Y_r] = \mathbb{E}[Y_1] + \cdots + \mathbb{E}[Y_r] = \underbrace{1/p + \cdots + 1/p}_{\text{r times}} = r/p,$$

since each Y_i is a geometric random variable. ■

Proposition 12. *If $X \sim \mathrm{NB}(r,p)$, then $\mathrm{Var}(X) = r(1-p)/p^2$.*

This is r times the variance of a geometric random variable.

Proof. Since $Y_1,\ldots,Y_r$ **are independent**, we can write

$$\mathrm{Var}(X) = \mathrm{Var}(Y_1 + \cdots + Y_r) = \mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_r) = \frac{r(1-p)}{p^2},$$

since each Y_i is a geometric random variable. ■

Poisson Distributions

Definition 13. The *Poisson distribution* with parameter λ is the discrete probability distribution of the number of events occurring in a fixed time interval assuming that

- the number of events occur at a fixed rate λ and
- the number of occurrences in nonoverlapping intervals are independent.

If X is a random variable that follows the Poisson distribution with parameter λ, we write $X \sim \mathrm{Pois}(\lambda)$, and its pmf is of the form for $k \geq 0$

Note here we are defining the pmf.

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

Proposition 14. *If $X \sim \mathrm{Pois}(\lambda)$, then $\mathbb{E}[X] = \lambda$.*

Proof. By definition,

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

by the Taylor series for e^x. ■

Recall that the Taylor series for e^x is $e^x = \sum_{k=0}^{\infty} x^k/k!$.

Proposition 15. *If* $X \sim \text{Pois}(\lambda)$, *then* $\text{Var}(X) = \lambda$.

Proof. We calculate $\mathbb{E}[X(X-1)]$. By linearity of expectation, we have

$$\mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \text{Var}(X).$$

The calculation is

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

Then $\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$. ■

It turns out that a Poisson random variable can be approximated by a binomial random variable. If λ is the number of successes in a fixed time period of length 1, then we may divide this into n time periods of length $1/n$. Each subinterval has a probability of $\lambda \cdot (1/n) = \lambda/n$ of success. Therefore, $\text{Pois}(\lambda) \approx \text{Bin}(n, \lambda/n)$, and the approximation gets better as $n \to \infty$ (or equivalently as $h \to 0$). To make this precise, note that for any $k \geq 0$ we have

$$\begin{aligned}\lim_{n\to\infty} P(\text{Bin}(n, \lambda/n) = k) &= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-(k-1))}{n^k}\right) \\ &\quad \times \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-(k-1)}{n}\right) \\ &\quad \times \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right) \\ &\quad \times \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 \\ &= \frac{\lambda^k e^{-\lambda}}{k!} \\ &= P(\text{Pois}(\lambda) = k).\end{aligned}$$

Exercises

Exercise 1. A basketball player makes each free throw with 9/10 probability. Suppose the player attempts 20 free throws.

(i) What is the probability that the player makes all 20 free throws?

(ii) What is the probability that the player makes 15 free throws?

(iii) What is the probability that the player makes at least 10 free throws?

(iv) What is the number of free throws the player is expected to make?

(v) What is the variance in the number of free throws the player makes?

Exercise 2. A factory produces items independently, each of which has probability $1/4$ of being defective. A batch contains 20 items.

(i) Find the probability that at least 3 items are defective and at most 6 are defective.

(ii) Given that at least one item is defective, find the probability exactly two are defective. There should be no summation $\sum$ in your answer.

(iii) Suppose the batch is accepted if at most 2 items are defective. Find the probability that the batch is accepted.

(iv) The factory inspects items one at a time until it finds the first defective item. Let N denote the number of items inspected, including the first defective item. Find $P(N = 5)$. Explain why N does not follow a binomial distribution. What distribution does it follow?

(v) Now suppose the defect probability p is unknown. You observe that exactly 4 of the 20 items are defective. Find the value of p that maximizes the likelihood of this observation.

Exercise 3. You flip a weighted coin that lands on heads with probability $1/5$ and on tails with probability $4/5$.

(i) What is the probability that you see your first heads at the 6th flip?

(ii) Find the probability that you flip the coin at least 5 times before getting heads.

(iii) Find the expected number of attempts you need to flip your coin until you get heads.

(iv) What is the variance in the number of attempts you need to flip until you get heads?

(v) Given that the first two flips were tails, what is the probability that the first heads is on the 5th flip?

Exercise 4. You flip a weighted coin that lands on heads with probability 1/3 and on tails with probability 2/3.

(i) Find the probability that the third head appears on the 7th flip.

(ii) Find the probability that at least 7 flips are needed to obtain the third head.

(iii) Find the expected number of attempts you need to flip your coin until you get three heads.

(iv) What is the variance in the number of attempts you need to flip until you get three heads.

(v) Given that the first two flips were tails, what is the probability that the third heads is on the 7th flip?

Exercise 5. A university help desk receives an average of 4 support requests per hour. Assume requests arrive according to a Poisson process, with a constant arrival rate and independent numbers of arrivals in disjoint time intervals.

(i) Find the probability that the help desk receives exactly 6 requests in one hour.

(ii) Find the probability that it receives at most 2 requests in one hour.

(iii) Given that 3 requests arrived in the first hour, find the probability that exactly 2 requests arrive in the next hour.

(iv) Find the expectation and variance of the number of requests received in a chosen hour.

Continuous Distributions

Now we shift our focus to continuous distributions. As the name suggests, this distribution describes random variables that are *continuous*. For example, if we want to analyze a random number between a and b, the random variable is

$$X = \text{number obtained between } a \text{ and } b.$$

In fact, there is no way to program a computer to output a random number, because such an algorithm is by contruction *deterministic*. Talk to me for more details.

Of course, X can behave differently according to the distribution of X. Now there is no notion of a pmf, but there is a notion of a *probability density function* (or pdf). This is motivated by the physical argument that if we integrate density across a dimension we obtain the mass.

- If we want our X to be uniformly distributed or have a *uniform distribution*, we want our pdf to be

  $$f(x) = \frac{1}{b-a}$$

  for all $x \in [a, b]$.

- Of course, X can have different distributions.

Definition 16. The *probability density function* (pdf) of a continuous random variable X is a function $f(x)$ that satisfies

Notice that these are exactly the axioms of probability.

- $f(x) > 0$ for all $x \in S$,
- $\int_S f(x)\,dx = 1$, and
- $P(X \in (a, b)) = \int_a^b f(x)\,dx$ for all $(a, b) \subseteq S$.

Compare this to the definition of a pmf of a discrete random variable, and how we replace sums with integrals.

Definition 17. The *probability mass function* (pmf) of a discrete random variable X is a function $f(x)$ that satisfies

Notice that these are exactly the axioms of probability.

- $f(x) > 0$ for all $x \in S$,
- $\sum_{x \in S} f(x) = 1$, and
- $P(X \in A) = \sum_{x \in A} f(x)$ for all $A \subseteq S$.

Similarly, the definitions of expected value and variance change analogously for continuous random variables.

Definition 18. If X is a continuous random variable with pdf f over a sample space $[a, b]$, then its *expected value* is

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

Definition 19. If X is a continuous random variable with pdf f over a sample space $[a, b]$, then its *variance* is

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \int_a^b (x - \mathbb{E}[X])^2 f(x)\, dx.$$

Compare this to the definitions of a discrete random variable.

Definition 20. If X is a continuous random variable with pmf f over a sample space S, then its *expected value* is

$$\mathbb{E}[X] = \sum_{x \in S} x f(x).$$

Definition 21. If X is a continuous random variable with pmf f over a sample space $[a, b]$, then its *variance* is

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \sum_{x \in S} (x - \mathbb{E}[X])^2 f(x).$$

Exercise 6. A bus arrives at a stop at a random time uniformly distributed between 0 and 20 minutes after you arrive. Let X be your waiting time in minutes.

(a) Find the probability density function of X.

(b) Find $\mathbb{E}[X]$.

(c) Find $\mathrm{Var}(X)$.

(d) Find the probability that you wait more than 15 minutes.

Exercise 7. A particle is confined to a one-dimensional infinite potential well on the interval $[0, L]$. Suppose its normalized wave function in the ground state is

$$\psi(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{\pi x}{L}\right)$$

for $x \in [0, L]$. According to Born's rule, the probability density of finding the particle at position x is $|\psi(x)|^2$.

(a) Find the probability density function $f(x)$ for the particle's position at x.

(b) Find the cumulative distribution function.

(c) Find the probability that the particle is found in the left half of the well.

(d) Find the probability that the particle is found in the middle third of the well.

(e) Find the probability that the particle is found within a distance $L/4$ of the center of the well.

(f) Find the expected position of the particle and its variance.