Lecture 05: Discrete Distributions

Mathematical Statistics I, MATH 60061/70061

Tuesday September 14, 2021

Reference: Casella & Berger, 3.1-3.2

Independence of two random variables

Random variables X and Y are said to be **independent** if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$

for all $x, y \in \mathbb{R}$.

In the discrete case, this is equivalent to the condition

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

for all x, y with x in the support of X and y in the support of Y.

If X and Y are independent, then any function of X is independent of any function of Y.

Independence of many random variables

Random variables $X_1, \ldots, X_n$ are independent if

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \ldots P(X_n \leq x_n),$$

for all $x_1, \ldots, x_n \in \mathbb{R}$.

Independence of many random variables

Random variables $X_1, \ldots, X_n$ are independent if

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \ldots P(X_n \leq x_n),$$

for all $x_1, \ldots, x_n \in \mathbb{R}$.

The criteria for independence of n events requires pairwise independence for all $\binom{n}{2}$ pairs, three-way independence for all $\binom{n}{3}$ triplets, and so on.

Here we need only one equality (?)

Independence of many random variables

Random variables $X_1, \ldots, X_n$ are independent if

$$P(X_1 \le x_1, \ldots, X_n \le x_n) = P(X_1 \le x_1) \ldots P(X_n \le x_n),$$

for all $x_1, \ldots, x_n \in \mathbb{R}$.

The criteria for independence of n events requires pairwise independence for all $\binom{n}{2}$ pairs, three-way independence for all $\binom{n}{3}$ triplets, and so on.

Here we need only one equality (?)

- We need the equality to hold for *all* possible $x_1, \ldots, x_n \in \mathbb{R}$.
- This requires *all* subset of the random variables to be independent.
- E.g., to verify pairwise independence between X_i and X_j, let all the x_k other than x_i, x_j go to ∞.

Properties of variance

For any random variable X and any constant c,

$$\begin{aligned}\mathrm{Var}(X+c) &= \mathrm{Var}(X),\\ \mathrm{Var}(cX) &= c^2\mathrm{Var}(X).\end{aligned}$$

If two random variables X and Y are *independent*, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Variance is the expectation of the nonnegative random variable $(X - EX)^2$, so $\mathrm{Var}(X) \geq 0$, with equality if and only if $P(X = a) = 1$ for some constant a.

Variance is *not* linear.

Bernoulli distribution

A random variable X is said to have the **Bernoulli distribution** with **parameter** p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$.

- There is not just one Bernoulli distribution, but rather a *family* of Bernoulli distributions, indexed by p.
- Any random variable whose possible values are 0 and 1 has a $\text{Bern}(p)$ distribution, where p is the probability of the random variable equaling 1.

Bernoulli trial

An experiment that can result in either a "success" or a "failure" (but not both) is called a **Bernoulli trial**.

A Bernoulli random variable can be thought of as the *indicator of success* in a Bernoulli trial. It equals

- 1 if success occurs in the trial
- 0 if failure occurs in the trial

With this interpretation, the parameter p is often called the *success probability* of the $\mathrm{Bern}(p)$ distribution.

Binomial distribution

Suppose that n *independent* Bernoulli trials are performed, each with the same success probability p.

Let X be the number of successes. The distribution of X is called the **Binomial distribution** with parameters n and p:

$$X \sim \mathrm{Bin}(n, p),$$

where n is a positive integer and $0 < p < 1$.

Binomial PMF

If $X \sim \text{Bin}(n, p)$, then the PMF of X is

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

for $x = 0, 1, \ldots, n$, and $P(X = x) = 0$ otherwise.

An experiment consisting of n independent Bernoulli trials produces a sequence of successes and failure.

- $P(\text{any sequence of } x \text{ successes}) = p^x (1-p)^{n-x}$
- There are $\binom{n}{x}$ such sequences
- Therefore, $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \ldots, n$

Binomial as the # of successes in n Bernoulli trials

If $X \sim \text{Bin}(n, p)$, viewed as the # of successes in n independent Bernoulli trials with success probability p, then we can write

$$X = X_1 + \cdots + X_n,$$

where the X_i are independent and identically distributed (i.i.d.) $\text{Bern}(p)$.

Binomial as the # of successes in n Bernoulli trials

If $X \sim \text{Bin}(n, p)$, viewed as the # of successes in n independent Bernoulli trials with success probability p, then we can write

$$X = X_1 + \cdots + X_n,$$

where the X_i are independent and identically distributed (i.i.d.) $\text{Bern}(p)$.

If $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, and X is independent of Y, then $X + Y \sim \text{Bin}(n + m, p)$.

- X is the sum of n i.i.d. $\text{Bern}(p)$: $X = X_1 + \cdots + X_n$
- Y is the sum of m i.i.d. $\text{Bern}(p)$: $Y = Y_1 + \cdots + Y_m$

$X + Y$ is the sum of $n + m$ i.i.d. $\text{Bern}(p)$, so its distribution is $\text{Bin}(n + m, p)$.

Sum of independent Binomial random variables (LOTP)

If $X \sim \mathrm{Bin}(n, p)$, $Y \sim \mathrm{Bin}(m, p)$, and X is independent of Y, then $X + Y \sim \mathrm{Bin}(n + m, p)$.

$$
\begin{aligned}
P(X+Y=k) &= \sum_{j=0}^{k} P(X+Y=k \mid X=j)P(X=j) \quad [\text{LOTP}] \\
&= \sum_{j=0}^{k} P(Y=k-j)P(X=j) \quad [X, Y \text{ independent}] \\
&= \sum_{j=0}^{k} \binom{m}{k-j} p^{k-j} q^{m-k+j} \binom{n}{j} p^j q^{n-j} \\
&= p^k q^{n+m-k} \sum_{j=0}^{k} \binom{m}{k-j} \binom{n}{j} \\
&= \binom{n+m}{k} p^k q^{n+m-k} \quad [\text{Vandermonde's identity}]
\end{aligned}
$$

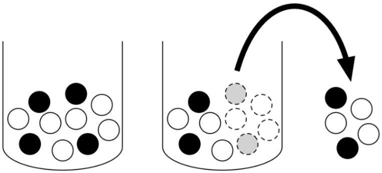
Drawing balls

Consider an urn with w white balls and b black balls, and draw n balls out of the urn:

- When sampling *with replacement*, the # of white balls follows a *Binomial distribution*

$$X \sim \mathrm{Bin}(n, w/(w+b))$$

- When sampling *without replacement*, the # of white balls follows a *Hypergeometric distribution*



Hypergeometric PMF

If $X \sim \mathrm{HGeom}(w, b, n)$, then the PMF of X is

$$P(X = x) = \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}},$$

for integers x satisfying $0 \leq x \leq w$ and $0 \leq n - x \leq b$, and $P(X = x) = 0$ otherwise.

The story of the Hypergeometric with w balls and b black balls

- $\binom{w+b}{n}$ ways to draw n balls out of $w + b$
- $\binom{w}{x}\binom{b}{n-x}$ ways to draw x white and $n - x$ black balls
- All samples are equally likely. So $P(X = x) = \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}}$

According to Vandermonde's identity, the PMF sums to 1.

Structure of the Hypergeometric story

Items in a population are classified using two sets of *tags*:

- Each ball is either white or black (the first set)
- Each ball is either sampled or not sampled (the second set)

At least one of these sets of tags is assigned completely at random.

Then, the number of twice-tagged items (e.g., balls are both white and sampled) follows a Hypergeometric distribution

$$X \sim \text{HGeom}(w, b, n).$$

Example: capture-recapture

Suppose there are N elk in a forest. Today, m of the elk are captured, tagged, and released back to the wild. At a later date, n elk are recaptured at random. Assume that an elk that has been captured does not learn how to avoid being captured again, so the recaptured elk are equally likely to be any set of n of the elk.

By the Hypergeometric story, the number of tagged elk in the recaptured sample is

$$X \sim \mathrm{HGeom}(m, N-m, n)$$

- An elk is either tagged or untagged (first set of tags)
- An elk is either captured or not captured at a later date (second set of tags)

Identical Hypergeometric distributions

If $X \sim \mathrm{HGeom}(w, b, n)$ and $Y \sim \mathrm{HGeom}(n, w + b - n, w)$, then X and Y have the same distribution.

Binomial vs. Hypergeometric

Consider an urn with w white balls and b black balls. Consider also drawing n balls as performing n Bernoulli trials:

- When sampling *with replacement*, the # of white balls follows a *Binomial distribution*

  $$X \sim \mathrm{Bin}(n, w/(w+b)).$$

  The Bernoulli trials involved are *independent*.

- When sampling *without replacement*, the # of white balls follows a *Hypergeometric distribution*

  $$X \sim \mathrm{HGeom}(w, b, n).$$

  The Bernoulli trials involved are *dependent*.

Binomial vs. Hypergeometric

Consider an urn with w white balls and b black balls. Consider also drawing n balls as performing n Bernoulli trials:

- When sampling *with replacement*, the # of white balls follows a *Binomial distribution*

$$X \sim \mathrm{Bin}(n, w/(w+b)).$$

  The Bernoulli trials involved are *independent*.

- When sampling *without replacement*, the # of white balls follows a *Hypergeometric distribution*

$$X \sim \mathrm{HGeom}(w, b, n).$$

  The Bernoulli trials involved are *dependent*.

What if $N = w + b \to \infty$?

Geometric distribution

Consider a sequence of independent Bernoulli trials, each with the same success probability $p \in (0, 1)$, with trials performed until a success occurs. Let X be the number of *failures* before the first successful trial. Then X has the **Geometric distribution** with parameter p, $X \sim \mathrm{Geom}(p)$.

The PMF of X is

$$P(X = x) = q^x p,$$

for $x = 0, 1, 2, \ldots$, where $q = 1 - p$.

This is a valid PMF (it sums to 1):

$$\sum_{x=0}^{\infty} q^x p = p \sum_{x=0}^{\infty} q^x = p \cdot \frac{1}{1-q} = 1.$$

Geometric expectation

Let $X \sim \mathrm{Geom}(p)$. By definition,

$$E(X) = \sum_{x=0}^{\infty} x q^x p,$$

where $q = 1 - p$.

- It is not a geometric series because of the extra x.
- Each term looks similar to xq^{x-1}, the derivative of q^x.

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q}$$

$$\Rightarrow \sum_{x=0}^{\infty} x q^{x-1} = \frac{1}{(1-q)^2}$$

$$\Rightarrow E(X) = \sum_{x=0}^{\infty} x q^x p = pq \frac{1}{(1-q)^2} = \frac{q}{p}.$$

Geometric variance

Taking another derivative:

$$\sum_{x=1}^{\infty} xq^x = \frac{q}{(1-q)^2} \Rightarrow \sum_{x=1}^{\infty} x^2 q^{x-1} = \frac{1+q}{(1-q)^3},$$

so

$$E(X^2) = \sum_{x=1}^{\infty} x^2 pq^x = pq\frac{1+q}{(1-q)^3} = \frac{q(1+q)}{p^2}.$$

Finally,

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = \frac{q(1+q)}{p^2} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}.$$

Negative Binomial distribution

In a sequence of independent Bernoulli trials with success probability $p \in (0, 1)$, if X is the number of *failures* before the rth success, then X is said to have the **Negative Binomial distribution** with parameters r and p, $X \sim \mathrm{NBin}(r, p)$.

Negative Binomial distribution

In a sequence of independent Bernoulli trials with success probability $p \in (0, 1)$, if X is the number of *failures* before the rth success, then X is said to have the **Negative Binomial distribution** with parameters r and p, $X \sim \mathrm{NBin}(r, p)$.

The PMF of X is

$$P(X = x) = \binom{x + r - 1}{r - 1} p^r q^x,$$

for $x = 0, 1, 2, \ldots$, where $q = 1 - p$.

Negative Binomial distribution

In a sequence of independent Bernoulli trials with success probability $p \in (0, 1)$, if X is the number of *failures* before the rth success, then X is said to have the **Negative Binomial distribution** with parameters r and p, $X \sim \mathrm{NBin}(r, p)$.

The PMF of X is

$$P(X = x) = \binom{x + r - 1}{r - 1} p^r q^x,$$

for $x = 0, 1, 2, \ldots$, where $q = 1 - p$.

- Imagine a string of 0's and 1's, with 1's representing successes.
- The probability of any specific string of x 0's and r 1's is $p^r q^x$.
- The string terminate in the rth 1; there are $(r - 1)$ 1's in the first $x + r - 1$ positions. So there are $\binom{x+r-1}{r-1}$ such strings.

Negative Binomial expectation

A Negative Binomial random variable $X \sim \mathrm{NBin}(r, p)$ can be viewed as the number of failures before the rth success in a sequence of independent Bernoulli trials with success probability p.

- X_1: the # of failures until the first success, $X_1 \sim \mathrm{Geom}(p)$
- X_2: the # of failures between the first success and the second success, $X_2 \sim \mathrm{Geom}(p)$
- $\ldots$
- X_r: the # of failures between the $(r-1)$th success and the rth success, $X_r \sim \mathrm{Geom}(p)$

So a Negative Binomial r.v. can be represented as a sum of i.i.d. Geometrics $X = X_1 + \cdots + X_r$, where the X_i are i.i.d. $\mathrm{Geom}(p)$.

Negative Binomial expectation and variance

A Negative Binomial random variable $X \sim \mathrm{NBin}(r, p)$ can be represented as a sum of i.i.d. Geometrics $X = X_1 + \cdots + X_r$, where the X_i are i.i.d. $\mathrm{Geom}(p)$.

By linearity, the expected value of X is

$$E(X) = E(X_1) + \cdots + E(X_r) = r \cdot \frac{q}{p}.$$

Since variance is additive for independent random variable,

$$\mathrm{Var}(X) = r \cdot \frac{q}{p^2}.$$

Poisson random variable

A random variable X has the Poisson distribution with parameter λ, where $\lambda > 0$, if the PMF of X is

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

for $x = 0, 1, 2, \ldots$.

The mean and variance of a $\mathrm{Pois}(\lambda)$ random variable are both equal to λ.

Poisson distribution

The Poisson distribution is often used in situations where we are counting the # of successes in a particular region or interval of time, and there are a large number of trials, each with a small probability of success.

- The # of emails you receive in an hour.
- The # of chips in a chocolate chip cookie.
- The # of earthquakes in a year in some region of the world.

The parameter λ is interpreted as the *rate* of occurrence of these rare events.

In the above examples, λ could be 20 (emails per hour), 10 (chips per cookie), and 2 (earthquakes per year).

Poisson paradigm

Let $A_1, \ldots, A_n$ be events with $p_j = P(A_j)$, where n is large, the p_j are small, and the A_j are independent or weakly dependent. Let

$$X = \sum_{j=1}^{n} I(A_j)$$

count how many of the A_j occur. Then X is approximately distributed as $\mathrm{Pois}(\lambda)$, with $\lambda = \sum_{j=1}^{n} p_j$.

The Poisson paradigm is also called the *law of rare events*, where the interpretation of "rare" is that the p_j are small, not that λ is small.

Sum of independent Poissons

If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$.

Sum of independent Poissons

If $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$, and X is independent of Y, then $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.

$$\begin{aligned}
P(X+Y=k) &= \sum_{j=0}^{k} P(X+Y=k \mid X=j)P(X=j) \quad \text{[LOTP]} \\
&= \sum_{j=0}^{k} P(Y=k-j)P(X=j) \\
&= \sum_{j=0}^{k} \frac{e^{-\lambda_2}\lambda_2^{k-j}}{(k-j)!}\frac{e^{-\lambda_1}\lambda_1^{j}}{j!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\binom{k}{j}\lambda_1^j\lambda_2^{k-j} \qquad \text{[Binomial theorem]} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^k}{k!}.
\end{aligned}$$

Poisson given a sum of Poissons

If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then the conditional distribution of X given $X + Y = n$ is $\mathrm{Bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.

Poisson given a sum of Poissons

If $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$, and X is independent of Y, then the conditional distribution of X given $X + Y = n$ is $\text{Bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.

$$
\begin{aligned}
P(X = k \mid X + Y = n) &= \frac{P(X + Y = n \mid X = k)P(X = k)}{P(X + Y = n)} \\
&= \frac{P(Y = n - k)P(X = k)}{P(X + Y = n)} \\
&= \frac{\left(\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right)\left(\frac{e^{-\lambda_1}\lambda_1^{k}}{k!}\right)}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} \\
&= \binom{n}{k}\frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.
\end{aligned}
$$

Poisson approximation to Binomial

If $X \sim \mathrm{Bin}(n, p)$ and we let $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains fixed, the the PMF of X converges to the $\mathrm{Pois}(\lambda)$ PMF.

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

Letting $n \to \infty$ with k fixed,

$$\frac{n(n-1)\dots(n-k+1)}{n^k} \to 1,$$

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda},$$

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1,$$

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

Example: visitors to a website

The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good approximation for the probability of getting *at least three* visitors on a particular day.

Example: visitors to a website

The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good approximation for the probability of getting *at least three* visitors on a particular day.

Binomial solution: Let $X \sim \mathrm{Bin}(n, p)$ be the number of visitors, where $n = 10^6$

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \binom{n}{0}(1-p)^n - \binom{n}{1}p(1-p)^{n-1} - \binom{n}{2}p^2(1-p)^{n-2}. \end{aligned}$$

Example: visitors to a website

The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good approximation for the probability of getting *at least three* visitors on a particular day.

Binomial solution: Let $X \sim \mathrm{Bin}(n, p)$ be the number of visitors, where $n = 10^6$

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \binom{n}{0}(1-p)^n - \binom{n}{1}p(1-p)^{n-1} - \binom{n}{2}p^2(1-p)^{n-2}. \end{aligned}$$

Since n is so large and p is so small, it is easy to run into computational difficulties in exact calculations.

Example: visitors to a website, continued

The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good approximation for the probability of getting *at least three* visitors on a particular day.

Since n is large, p is small, and $np = 2$ is moderate, $\mathrm{Pois}(2)$ is a good approximation. This gives

$$P(X \geq 3) = 1 - P(X < 3) \approx 1 - e^{-2} - e^{-2} \cdot 2 - e^{-2} \cdot \frac{2^2}{2!} \approx 0.3233.$$