W04 Notes

Bernoulli process

01 Theory - Bernoulli, binomial, geometric, Pascal, uniform

In a Bernoulli process, an experiment with binary outcomes is repeated; for example flipping a coin repeatedly. Several discrete random variables may be defined in the context of some Bernoulli process.

Notice that the sample space of a Bernoulli process is infinite: an outcome is any sequence of trial outcomes, e.g. $HTHHTTHHHTTTHHHHTTTT\cdots$

Bernoulli variable

A random variable X_i is a **Bernoulli indicator**, written $X_i \sim \text{Ber}(p)$, when X_i indicates whether a success event, having probability p, took place in trial number i of a Bernoulli process.

Bernoulli PMF:

$$P_{X_i}(k) \;=\; \begin{cases} p & k = 1 \\ q & k = 0 \\ 0 & \text{else} \end{cases}$$

Here $q = 1 - p$.

An RV that always gives either 0 or 1 for every outcome is called an **indicator variable**.

Binomial variable

A random variable S is **binomial**, written $S \sim \text{Bin}(n,\, p)$, when S counts the *number of successes* in a Bernoulli process, each having probability p, over a specified number n of trials.

Binomial PMF:

$$P_S(k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } \; k = 0, 1, 2, \ldots, n$$

- For example, if $S \sim \text{Bin}(10,\, 0.2)$, then $P_S(5)$ gives the odds that success happens exactly 5 times over 10 trials, with probability 0.2 of success for each trial.
- In terms of the Bernoulli indicators, we have: $S = X_1 + X_2 + \cdots + X_n$
- If A is the success event, then $p = P[A]$ is the success probability, and $q = 1 - p$ is the failure probability.

Geometric variable

A random variable N is **geometric**, written $N \sim \text{Geom}(p)$, when N counts the *discrete wait time* in a Bernoulli process until the *first success* takes place, given that success has probability p in each trial.

Geometric PMF:

$$P_N(k) \;=\; q^{k-1}p \qquad \text{for } k = 1, 2, 3, \ldots$$

Here $q = 1 - p$.

- For example, if $N \sim \text{Geom}(30\%)$, then $P_N(7)$ gives the probability of getting: failure on the first 6 trials AND success on the 7^{th} trial.

Pascal variable

A random variable L is **Pascal**, written $L \sim \text{Pasc}(\ell,\, p)$, when L counts the *discrete wait time* in a Bernoulli process until success happens *ℓ times*, given that success has probability p in each trial.

Pascal PMF:

$$P_L(k) \;=\; \binom{k-1}{\ell-1}(1-p)^{k-\ell}p^{\ell} \qquad \text{for } k = \ell, \ell+1, \ell+2, \ldots$$

- For example, if $L \sim \text{Pasc}(3,\, 0.25)$, then $P_L(8)$ gives the probability of getting: the 3^{rd} success on (precisely) the 8^{th} trial.
- Interpret the formula: # ways to arrange 2 successes among 7 'prior' trials, times the probability of exactly 3 successes and 5 failures in one specific sequence.
- The Pascal distribution is also called the **negative binomial** distribution, e.g. $L \sim \text{Negbin}(\ell,\, p)$.

Uniform variable

A discrete random variable X is **uniform** on a finite set $A \subset S$, written $X \sim \text{Unif(A)}$, when the probability is a fixed constant for outcomes in A and zero for outcomes outside A.

Discrete uniform PMF:

$$P_X(k) = \begin{cases} \dfrac{1}{|A|} & \text{when } k \in A \\ 0 & \text{when } k \notin A \end{cases}$$

Continuous uniform PDF:

$$f_X(x) = \begin{cases} \frac{1}{P[A]} & \text{when } x \in A \\ 0 & \text{when } x \notin A \end{cases}$$

02 Illustration

Example - Roll die until

Roll a fair die repeatedly. Find the probabilities that:

(a) At most 2 threes occur in the first 5 rolls.

(b) There is no three in the first 4 rolls, using a geometric variable.

Solution

(a)

(1) Label variables and events:

Use a variable $S \sim \text{Bin}(5, 1/6)$ to count the number of threes among the first six rolls.

Seek $P[S \leq 2]$ as the answer.

---

(2) Calculations:

Divide into exclusive events:

$$P[S \leq 2] \quad \gg\gg \quad P_S(0) + P_S(1) + P_S(2)$$

$$\gg\gg \quad \binom{5}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 + \binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 + \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$$

$$\gg\gg \quad \frac{625}{648} \quad \gg\gg \quad \approx 0.965$$

(b)

(1) Label variables and events:

Use a variable $N \sim \text{Geom}(1/6)$ to give the roll number of the first time a three is rolled.

Seek $P[N > 4]$ as the answer.

---

(2) Compute:

Sum the PMF formula for Geom(1/6):

$$P[N > 4] \quad \ggg \quad \sum_{k=5}^{\infty} \left(\frac{5}{6}\right)^{k-1} \left(\frac{1}{6}\right)$$

---

(3) Recall geometric series formula:

For any geometric series:

$$a + ar + ar^2 + ar^3 + \cdots \quad = \quad \frac{a}{1-r}$$

Therefore:

$$P[N > 4] \quad = \quad \sum_{k=5}^{\infty} \left(\frac{5}{6}\right)^{k-1} \left(\frac{1}{6}\right) \quad \ggg \quad \left(\frac{5}{6}\right)^4$$

Example - Cubs winning the World Series

Suppose the Cubs are playing the Yankees for the World Series. The first team to 4 wins in 7 games wins the series. What is the probability that the Cubs win the series?

Assume that for any given game the probability of the Cubs winning is $p = 45\%$ and losing is $q = 55\%$.

Solution

Method (a): We solve the problem using a binomial distribution.

(1) Label variables and events:

Use a variable $X \sim \text{Bin}(7, p)$. This X counts the number of wins over 7 games. Thus, for example, $P_X(4)$ is the probability that the Cubs win exactly 4 games over 7 played.

Seek $P_X(4) + P_X(5) + P_X(6) + P_X(7)$ as the answer.

---

(2) Calculate using binomial PMF:

$$P_X(k) = \binom{7}{k} p^k q^{7-k}$$

Insert data:

$$P_X(4) + \cdots + P_X(7)$$

$$\ggg \quad \binom{7}{4} p^4 q^3 + \binom{7}{5} p^5 q^2 + \binom{7}{6} p^6 q^1 + \binom{7}{7} p^7 q^0$$

Compute:

$$\gg\gg \quad \frac{7\cdot 6\cdot 5}{3\cdot 2}p^4q^3+\frac{7\cdot 6}{2}p^5q^2+\frac{7}{1}p^6q^1+1\cdot p^7q^0$$

$$\gg\gg \quad p^4\left(35q^3+21p^1q^2+7p^2q+p^3\right)$$

Convert $q \gg (1-p)$:

$$\gg\gg \quad p^4\left(35(1-p)^3+21p(1-p)^2+7p^2(1-p)+p^3\right)$$

$$\gg\gg \quad 35p^4-84p^5+70p^6-20p^7 \quad \gg\gg \quad \approx 0.39$$

Method (b): We solve the problem using a Pascal distribution instead.

(1) Label variables and events:

Use a variable $Y \sim \text{Pasc}(4,\, p)$. This Y measures the discrete wait time until the 4^{th} win. Thus, for example, $P_Y(k)$ is the probability that the Cubs win their 4^{th} game on game number k.

Seek $P_Y(4)+P_Y(5)+P_Y(6)+P_Y(7)$ as the answer.

---

(2) Calculate using Pascal PMF:

$$P_Y(k)=\binom{k-1}{3}p^4q^{k-4}$$

Insert data:

$$P_Y(4)+\cdots+P_Y(7)$$

$$\gg\gg \quad \binom{3}{3}p^4q^0+\binom{4}{3}p^4q^1+\binom{5}{3}p^4q^2+\binom{6}{3}p^4q^3$$

Compute:

$$\gg\gg \quad 1\cdot p^4+\frac{4}{1}\cdot p^4q^1+\frac{5\cdot 4}{2}p^4q^2+\frac{6\cdot 5\cdot 4}{3\cdot 2}p^4q^3$$

$$\gg\gg \quad p^4\left(1+4q+10q^2+20q^3\right)$$

Convert $q \gg (1-p)$:

$$\gg\gg \quad p^4\left(1+4(1-p)+10(1-p)^2+20(1-p)^3\right)$$

$$\gg\gg \quad 35p^4-84p^5+70p^6-20p^7 \quad \gg\gg \quad \approx 0.39$$

Notice: The calculation seems very different than method (a), right up to the end!

Expectation and variance

03 Theory - Expectation and variance

Expected value

The **expected value** $E[X]$ of random variable X is the weighted average of the values of X, weighted by the probability of those values.

Discrete formula using PMF:

$$E[X] = \sum_k k \cdot P_X(k)$$

Continuous formula using PDF:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x)\, dx$$

Notes:

- Expected value is sometimes called **expectation**, or even just **mean**, although the latter is best reserved for statistics.
- The Greek letter μ is also used in contexts where 'mean' is used.

Let X be a random variable, and write $E[X] = \mu$.

Variance

The **variance** $\text{Var}[X]$ measures the average *squared deviation* of X from μ. It estimates how *concentrated* X is around μ.

- Defining formula:

$$\text{Var}[X] = E[(X - \mu)^2]$$

- Shorter formula:

$$\text{Var}[X] = E[X^2] - E[X]^2$$

Calculating variance

- Discrete formula using PMF:

$$\text{Var}[X] = \sum_k (k - \mu)^2 P_X(k)$$

- Continuous formula using PDF:

$$\text{Var}[X] = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x)\, dx$$

Standard deviation

The quantity $\sigma_X = \sqrt{\text{Var}[X]}$ is called the **standard deviation** of X.

04 Illustration

Exercise - Tokens in bins

Consider a game like this: a coin is flipped; if H then draw a token from Bin 1, if T then from Bin 2.

- Bin 1 contents: 1 token \$1,000, and 9 tokens \$1
- Bin 2 contents: 5 tokens \$50, and 5 tokens \$1

It costs \$50 to enter the game. Should you play it? (A lot of times?) How much would you pay to play?

Solution >

(1) Setup:

Let X be a random variable measuring your winnings in the game.

The possible values of X are 1, 50, and 1000.

---

(2) Find PDF $P_X(k)$:

For $k = 1$ have $P_X(1) = \frac{1}{2} \cdot \frac{9}{10} + \frac{1}{2} \cdot \frac{5}{10} \quad \gg\gg \quad \frac{7}{10}$

For $k = 50$ have $P_X(50) = \frac{1}{2} \cdot \frac{5}{10} \quad \gg\gg \quad \frac{1}{4}$

For $k = 1000$ have $P_X(1000) = \frac{1}{2} \cdot \frac{1}{10} \quad \gg\gg \quad \frac{1}{20}$

These add to 1, and $P_X(x) = 0$ for all other x.

---

(3) Find $E[X]$ using the discrete formula:

$$E[X] = \sum_k k \cdot P_X(k) \quad \gg\gg \quad 1 \cdot P_X(1) + 50 \cdot P_X(50) + 1000 \cdot P_X(1000)$$

$$\gg\gg \quad 1 \cdot \frac{7}{10} + 50 \cdot \frac{1}{4} + 1000 \cdot \frac{1}{20} \quad \gg\gg \quad \approx 63.2$$

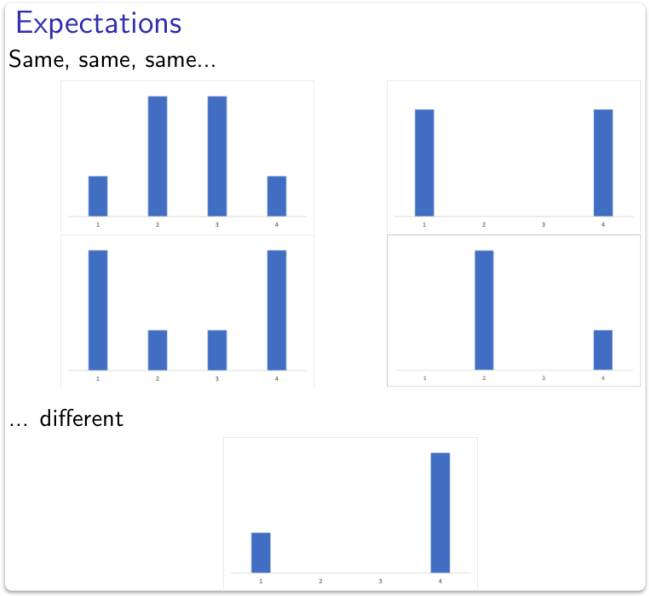
---

(4) Conclusion:

Since $63.2 > 50$, if you play it a lot at \$50 you will generally make money.

Challenge Q:

If you start with $200 and keep playing to infinity, how likely is it that you go broke?



Example - Expected value: rolling dice

Let X be a random variable counting the number of dots given by rolling a single die.

Then:

$$E[X] \quad \gg\gg \quad 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} \quad \gg\gg \quad \frac{7}{2}$$

Let S be an RV that counts the dots on a roll of *two* dice.

The PMF of S:

k	2	3	4	5	6	7	8	9	10	11	12
$p_S(k) = P(S = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Then:

$$E[S] \quad \gg\gg \quad 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \cdots + 12 \cdot \frac{1}{36} \quad \gg\gg \quad 7$$

Notice that $\frac{7}{2} + \frac{7}{2} = 7$.

In general, $E[X + Y] = E[X] + E[Y]$.

Let X be a green die and Y a red die.

From the earlier calculation, $E[X] = \frac{7}{2}$ and $E[Y] = \frac{7}{2}$.

Since $S = X + Y$, we derive $E[S] = 7$ by simple addition!

Example - Expected value by finding new PMF

Let X have distribution given by this PMF:

x	1	2	3	4	5
$p_X(x)$	$1/7$	$1/14$	$3/14$	$2/7$	$2/7$

Find $E[|X-2|]$.

Solution

(1) Compute the PMF of $|X-2|$.

PMF arranged by possible value:

$$\begin{array}{lll} P[|X-2|=0] & \gg\gg & P[X=2] = \frac{1}{14} \\ P[|X-2|=1] & \gg\gg & P[X=1] + P[X=3] = \frac{1}{7} + \frac{3}{14} = \frac{5}{14} \\ P[|X-2|=2] & \gg\gg & P[X=4] = \frac{2}{7} \\ P[|X-2|=3] & \gg\gg & P[X=5] = \frac{2}{7} \\ \vdots & & \vdots \\ P[|X-2|=k] & \gg\gg & 0 \quad \text{for } \ k \neq 0,1,2,3. \end{array}$$

---

(2) Calculate the expectation.

Using formula for discrete PMF:

$$E[|X-2|] = 0 \cdot \tfrac{1}{14} + 1 \cdot \tfrac{5}{14} + 2 \cdot \tfrac{2}{7} + 3 \cdot \tfrac{2}{7} = \tfrac{25}{14}$$

Exercise - Variance using simplified formula

Suppose X has this PMF:

k :	**1**	**2**	**3**
$P_X(k)$:	$1/7$	$2/7$	$4/7$

Find $\text{Var}[\frac{1}{1+X}]$ using the formula $\text{Var}[Y] = E[Y^2] - E[Y]^2$ with $Y = \frac{1}{1+X}$.

(Hint: you should find $E[Y] = \frac{13}{42}$ and $E[Y^2] = \frac{13}{126}$ along the way.)

Poisson process

05 Theory - Poisson variable

Poisson variable

A random variable X is **Poisson**, written $X \sim \text{Pois}(\lambda)$, when X counts the number of "arrivals" in a fixed "interval." It is applicable when:

- The arrivals come at a *constant average rate* λ.
- The arrivals are independent of each other.

Poisson PMF:

$$P_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

A Poisson variable is comparable with a binomial variable. Both count the occurrences of some "arrivals" over some "space of opportunity."

- The binomial opportunity is a set of *n repetitions* of a trial.
- The Poisson opportunity is a *continuous interval* of time.

In the binomial case, success occurs at some rate p, since $p = P[A]$ where A is the success event.

In the Poisson case, arrivals occur at some rate λ.

The Poisson distribution is actually the limit of binomial distributions by taking $n \to \infty$ while np remains fixed, so $p \to 0$ in perfect balance with $n \to \infty$.

Let $X_{n,p} \sim \text{Bin}(n, p)$ and let $Y_\lambda \sim \text{Pois}(\lambda)$. Fix λ and let $p = \lambda/n$. Then for any $k \in \mathbb{N}$:

$$P_{X_{n,p}}(k) \quad \overset{n\to\infty}{\longrightarrow} \quad P_{Y_\lambda}(k)$$

For example, let $X_{n,3/n} \sim \text{Bin}(n, 3/n)$, so $np = 3$, and look at $P_{X_{n,3/n}}(1)$ as $n \to \infty$:

$$P_{X_{n,3/n}}(1) \quad \gg\gg \quad \binom{n}{1}\left(\frac{3}{n}\right)^1\left(1-\frac{3}{n}\right)^{n-1}$$

$$\gg\gg \quad 3\left(1-\frac{3}{n}\right)^{n-1} \quad \longrightarrow \quad 3e^{-3} \quad \text{as } n \to \infty$$

Interpretation - Binomial model of rare events

Let us interpret the assumptions of this limit. For n large but p small such that $\lambda = np$ remains moderate, the binomial distribution describes a large number of trials, a low probability of success per trial, but a moderate total count of successes.

This setup describes a physical system with a large number of parts that may activate, but each part is unlikely to activate; and yet the number of parts is so large that the total number of arrivals is still moderate.

06 Illustration

Example - Radioactive decay is Poisson

Consider a macroscopic sample of Uranium.

Each atom decays independently of the others, and the likelihood of a single atom popping off is very low; but the product of this likelihood by the total number of atoms is a moderate number.

So there is some constant average rate of atoms in the sample popping off, and the number of pops per minute follows a Poisson distribution.

Example - Arrivals at a post office

Client arrivals at a post office are modelled well using a Poisson variable.

Each potential client has a very low and independent chance of coming to the post office, but there are many thousands of potential clients, so the arrivals at the office actually come in moderate number.

Suppose the average rate is 5 clients per hour.

(a) Find the probability that nobody comes in the first 10 minutes of opening. (The cashier is considering being late by 10 minutes to run an errand on the way to work.)

(b) Find the probability that 5 clients come in the first hour. (I.e. the average is achieved.)

(c) Find the probability that 9 clients come in the first two hours.

Solution

(a)

(1) Convert rate for desired window.

Expect $5/6$ clients every 10 minutes.

Let $X \sim \text{Pois}(5/6)$.

Seek $P_X(0)$ as the answer.

---

(2) Compute.

Formula:

$$P_X(k) = e^{-5/6}\frac{(5/6)^k}{k!}$$

Insert data and compute:

$$P_X(0) \quad \gg\gg \quad e^{-5/6} \quad \gg\gg \quad \approx 0.435$$

(b)

Rate is already correct.

Let $X \sim \text{Pois}(5)$.

Compute the answer:

$$P_X(5) = e^{-5}\frac{5^5}{5!} \quad \gg\gg \quad \approx 0.175$$

(c)

Convert rate for desired window.

Expect 10 clients every 2 hours.

Let $X \sim \text{Pois}(10)$.

Compute the answer:

$$P_X(9) \quad \gg\gg \quad e^{-10}\frac{10^9}{9!} \quad \gg\gg \quad \approx 0.125$$

Notice that 0.125 is smaller than 0.175.

07 Theory - Poisson limit of binomial

Extra - Derivation of binomial limit to Poisson

Consider a random variable $X \sim \text{Bin}(n, p)$, and suppose n is very large.

Suppose also that p is very small, such that $E[X] = np$ is *not* very large, but the extremes of n and p counteract each other. (Notice that then npq will *not* be large so the normal approximation does *not* apply.) In this case, the binomial PMF can be approximated using a factor of e^{-np}. Consider the following rearrangement of the binomial PMF:

$$P_X(k) \quad \gg\gg \quad \binom{n}{k}p^k q^{n-k}$$

$$\gg\gg \quad \frac{n(n-1)\cdots(n-k+1)}{k!}p^k(1-p)^n\frac{1}{q^k}$$

$$\gg\gg \quad (1-p)^n\frac{(np)^k}{k!}\left[\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdots\frac{n-k+1}{n}\right]\frac{1}{q^k}$$

Since n is very large, the factor in brackets is approximately 1, and since p is very small, the last factor of $1/q^k$ is also approximately 1 (provided we consider k small compared to n). So we have:

$$P_X(k) \approx (1-p)^n\frac{(np)^k}{k!}.$$

Write $\lambda = np$, a moderate number, to find:

$$P_X(k) \approx \left(1-\frac{\lambda}{n}\right)^n\frac{\lambda^k}{k!}.$$

Here at last we find $e^{-\lambda}$, since $\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$ as $n \to \infty$. So as $n \to \infty$:

$$P_X(k) \approx e^{-\lambda} \frac{\lambda^k}{k!}.$$

Extra - Binomial limit to Poisson and divisibility

Consider a sequence of increasing n with decreasing p such that $\lambda = np$ is held fixed. For example, let $n = 1, 2, 3, \ldots$ while $p = \frac{\lambda}{n}$.

Think of this process as increasing the number of causal agents represented: group the agents together into n bunches, and consider the odds that such a bunch activates. (For the call center, a bunch is a group of users; for radioactive decay, a bunch is a unit of mass of Uranium atoms.)

As n doubles, we imagine the number of agents per bunch to drop by half. (For the call center, we divide a group in half, so twice as many groups but half the odds of one group making a call; for the Uranium, we divide a chunk of mass in half, getting twice as many portions with half the odds of a decay occurring in each portion.

This process is formally called *division of a distribution*, and the fact that the Poisson distribution arises as the limit of such division means that it is infinitely divisible.

Extra - Theorem: Poisson approximation of the binomial

Suppose $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Pois}(np)$. Then:

$$\left|P_X(k) - P_Y(k)\right| \leq np^2$$

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

In consequence of this theorem, a Poisson distribution may be used to approximate the probabilities of a binomial distribution for large n when it is impracticable (even for a computer) to calculate large binomial coefficients.

The theorem shows that the Poisson approximation is appropriate when np is a moderate number while np^2 is a small number.